

Don't mention the 'F' word

The National Institutes of Health's plan for obesity research is undermined by the refusal of the Bush administration to commit to regulation. More attention needs to be paid to public health and less to the interests of the food industry.

"Abountiful food supply with abundant choices of relatively inexpensive, calorically-dense food products that are convenient and taste good." If this phrase conjures up images of fresh fruit, fish and coconuts, think again. It refers to an American diet including many calorie-packed convenience foods, soft drinks and snacks, but in the sanitized parlance of the new plan for obesity research from the US National Institutes of Health (NIH). This bountiful supply, it tells us, is to blame for "the current 'obesogenic' environment that promotes increased caloric intake".

Such vernacular fits a consistent trend across the Bush administration's so-called anti-obesity policies: to tiptoe around calling a hamburger a hamburger, and state unequivocally that many of the foods marketed to the public are the biggest single cause of the obesity epidemic. If government or the NIH tells people to eat less, they must then say less of what, and will be in trouble with the powerful food industry. The administration's new anti-obesity strategy (see page 244) is a series of half-measures, doling out lessons to eat better and exercise more, dressed up as personal freedom and choice. It blatantly shirks committing the government to regulating the food industry.

Individual choice is just not working (see page 252). Obesity is

spiralling out of control, with 130 million Americans, or 64%, overweight or obese. They are desperate for help, spending \$37.1 billion annually on diet books and fads offering instant remedies. Bold government intervention, including compulsory calorie labels and a clampdown on junk-food marketing, is needed.

"We're just too darn fat, ladies and gentlemen, and we are going to do something about it," Tommy Thompson, head of the US Department of Health and Human Services, told a press conference last week to launch the new strategy. "We need to tackle America's weight issues as aggressively as we are addressing smoking and tobacco." Fighting talk, in contrast with his department's efforts to dilute the World Health Organization's Global Strategy on Diet, Physical Activity and Health, which highlights the food industry's role in reducing obesity.

Americans deserve better. The medical and other costs of obesity to the United States total more than \$117 billion a year. With obesity overtaking tobacco as the main preventable cause of death, this is an issue of national importance. Scientists and others now need to make their voices heard, and to call on the candidates in the forthcoming US presidential elections to clearly state that they will stop tiptoeing around the elephant in the room in the obesity debate. ■

Making data dreams come true

If new bioinformatics initiatives deliver, cancer researchers can expect a gradual revolution in working practice.

Imagine that for selected cancer patients, biopsies are taken before, during and after treatment, made anonymous and the analyses stored promptly in an accessible fashion. These biopsy samples are subjected to gene-expression and proteomic analysis, and these molecular data can also be stored accessibly. Imagine also that the patient's data can readily be compared with those from other trials. And imagine that one can drill down into clinical and other databases in an intelligent search in hours rather than months. One end-point might be the rapid identification of individualized molecular profiles correlated with sensitivity or resistance to therapy.

This vision requires common standards of data storage at each level of investigation, new frameworks for cross-referencing terms and their biological contexts ('ontologies') between disparate types of data, and new bioinformatics tools to make it all practicable. The benefits? Quicker routes to identifying patients' individual characteristics that make one treatment more appropriate than another; easier integration of genomics research into clinical trials; and much readier access by basic molecular and cell biologists to the early lessons that can be drawn from even a few patients, as well as from large-scale, randomized clinical trials.

Much of this is beginning to be realized. The US National Cancer Institute last week launched its Cancer Biomedical Informatics Grid project, bringing together the institute's Center for Bioinformatics (NCICB) with translational centres and clinical-trial cooperative cancer centres in the United States (see <http://cabig.nci.nih.gov>). The programme, which costs \$20 million a year, will yield networks of clinical-trial information, tissue data, ontological development

and integrative tools that give researchers ready access to data.

In close collaboration, Britain is embarking on the same route. The strategic body that represents most of the relevant UK funding agencies, the National Cancer Research Institute (NCRI), is this week announcing in *Nature* (see page xii and also www.cancerinformatics.org.uk) a framework for similar developments*.

Although the NCRI efforts will involve centres of excellence, they currently lack the dedicated budget of the NCICB. It is therefore critical that funding agencies collectively set aside millions each year for these developments, and that a combination of strategic leadership and peer review is established, as in the US programme.

One unique part of the UK landscape is both a boon and a potential weakness: the National Health Service (NHS). Its existence gives Britain a greater ability to focus its efforts on standards and storage than the patchwork of state and privately funded health services in the United States allows. But information technology in the NHS is in dire need of development. And whatever happens, it is essential that suppliers of the technology do not gain control over access to public data, as has happened in other contexts.

More positively, despite the need of researchers to protect not only their patients but also their competitive interests, the leaders of these bioinformatics initiatives have been gratified by the positive attitudes to data sharing encountered so far on both sides of the Atlantic. Next, and sooner rather than later, comes the challenge of extending cancer bioinformatics collaboration across the disparate research and health systems of Europe. ■

*The framework's steering panel includes the Editor of *Nature*.





First base
Genome offers hints
on why chimps and
humans are different
p242



Body of evidence
University cancels
donor programme
after arrests
p243



Positive action
Pentagon seeks out
the secrets of negative
refraction
p245



Sinking pints
Proof that stout
bubbles take the
plunge
p247

Battle lines are drawn as French researchers resign en masse

Declan Butler, Paris

A thundering roar went up from thousands of researchers gathered outside the city hall in Paris on 9 March. Leaders of a scientific revolt emerged to announce that more than 2,000 lab heads meeting inside had voted overwhelmingly to resign in protest at the government's handling of research.

The news was a massive boost for the protest movement Save Research. The 72,000 signatories of the group's petition, launched in January, had promised to resign from any management duties unless the government met their demands. These included redressing a financial crisis in research, reinstating 550 full-time research posts that have been switched to short-term contracts, and launching a national debate on reforming French science (see *Nature* 428, 105; 2004).

The vote to carry through the threat showed that senior members of the scientific establishment were prepared to act. "The struggle continues," says an exhilarated Alain Trautmann, a cell biologist at the Cochin Institute in Paris and spokesman for Save Research.

Resolve had hardened on the morning of the vote, when Prime Minister Jean-Pierre Raffarin gave a hardline interview with the newspaper *Liberation* in which he insisted that the government would not meet the protesters' demands. The subsequent show of strength far exceeded the protest organizers' expectations. Once fax and e-mail votes were counted, 1,280 lab heads and 1,958 team leaders had resigned. This amounts to more than half of the lab heads at both the CNRS, the country's largest research agency, and the biomedical research agency INSERM.

The move is largely symbolic in the short term, as the research-agency heads have two months to consider whether to accept the resignations. Bernard Larrouturou, CNRS director-general, has already convened a meeting with the protesters for 23 March.

In the meantime, researchers are setting up local committees to maintain the pressure, and Save Research has called for a national protest on 19 March. It has also asked young scientists and homesick French expatriates to inundate the research ministry with their CVs.

The resignations have opened up the



Solidarity: protesters rally on the streets of Paris as researchers vote to resign managerial duties.

debate about how to overhaul France's centralized and bureaucratic science system. The government was already planning to produce a white paper on the subject before the end of the year, but Save Research refused to cooperate with the consultation process, complaining that its views were not being represented.

This deadlock was broken last week when Save Research agreed to work with an independent committee set up to consult with French researchers and to feed their views into the white paper, and next year's science budget.

The 30-member committee, which will meet for the first time this week, will be co-chaired by Étienne-Émile Baulieu and Édouard Brézin, president and vice-president, respectively, of the French Academy of Sciences. It will include nine members of Save Research and other leading French scientists, and aims to organize a whirlwind series of regional and national conferences.

The committee is unlikely to have to look far for suggestions on reform. Larrouturou, for instance, last week discussed plans for the CNRS in the press. He wants to channel resources into a smaller number of larger labs, to boost the number of foreign researchers in

the agency's ranks from 12% to 25% and to make evaluation procedures tougher.

Meanwhile, four leading scientists set out their manifesto on 10 March. Biologist François Jacob and chemist Jean-Marie Lehn, Nobel laureates at the Collège de France in Paris, have worked with Philippe Kourilsky, director-general of the Pasteur Institute, and Pierre-Louis Lions, a mathematician at the University of Paris Dauphine, to draw up their unashamedly élitist vision. Importing many elements of the British and US research systems, their plan would decentralize the public research agencies, and over time devolve research to the French universities. "The crisis is open, and we must profit by reforming in depth," the manifesto's authors note.

More immediately, Save Research has called on President Jacques Chirac to intervene to create more posts for young scientists this year. Chirac has yet to reply, but with a recent poll suggesting that 82% of the French public support the scientists, the pressure is on for him to find a way out of the deadlock — especially in light of the regional elections that will be held at the end of March.

♦ <http://recherche-en-danger.apinc.org>

Geneticists study chimp–human divergence

Erika Check, San Diego

Scientists combing the draft of the chimpanzee genome sequence are finding tantalizing hints about the differences between humans and our closest relatives.

Results gleaned from the publicly available genome, which was released last December, were discussed for the first time at a workshop on 12 March at the University of California, San Diego.

Researchers led by Xose Puente at the University of Oviedo in Spain, for example, are working on genes that encode enzymes called proteases. These genes are almost identical in humans and chimps except for one subset, found in the immune system, that varies greatly between the two species.

This suggests proteases of the immune system diverged rapidly after chimps and humans became separate species. Such findings could help to clarify the differences between the two species. They could also explain why chimps are less severely affected than humans by AIDS or Alzheimer's disease, and might lead to treatments for such conditions.

Evan Eichler, a bioinformaticist at Case Western Reserve University in Cleveland, Ohio, found one section of DNA, 10,000 bases long, that is absent in the human genome but present in multiple copies in the



Relative merits: genome comparison may explain chimp resistance to AIDS.

chimpanzee and other African apes. By sequencing that section, Eichler's lab found that it was similar to a retrovirus that infects African monkeys. Because retroviruses insert themselves into host DNA, Eichler suspects that this virus might have jumped from monkeys into apes.

But experts say that such topics are hard to study, in part because there are still large gaps in the sequenced chimp genome. Reassuringly, Robert Waterston, a geneticist at the University of Washington in Seattle

who leads an international team studying the chimp genome, says that the US National Human Genome Research Institute will provide funding for sequencers to keep plugging the gaps.

Further help in studying chimp and human differences will come from molecular biologist Ajit Varki and computer scientist Chaitan Baru, both at the University of California, San Diego. They plan to collate details of all the characteristics that differ between chimps and humans on a web-based Museum of Comparative Anthropogeny (MOCA). They hope that MOCA will help to reveal how the 1% of the genome that differs between humans and chimps has translated into the huge differences between the two species in areas such as brain size and behaviour.

But biologist Caroline Tutin of the University of Stirling in Scotland warns that, although the differences are illuminating, geneticists should not overlook the similarities between humans and chimps. "Do not forget how similar the chimp and human are, because that is the basis for their conservation," Tutin says. She warns that as wild populations of endangered chimps dwindle, scientists risk losing a species that is not only a precious genetic resource, but a close relative as well. ■

Plague professor gets two years in bioterror case

Erika Check, Washington

Microbiologist Thomas Butler has been sentenced to two years in prison for convictions that stem from a January 2003 bioterrorism scare.

The case began when Butler reported vials of plague bacteria missing from his lab at Texas Tech University in Lubbock. After being questioned by the FBI, Butler said he might have accidentally destroyed the vials.

Investigators charged Butler with 69 crimes related to the incident, including lying to them and defrauding his university. Butler was cleared of the most serious charges last December, but was convicted of fraud against Texas Tech and of mislabelling a plague sample he shipped to Tanzania.

Butler's two-year sentence is more lenient than that sought by federal

prosecutors, who called for millions of dollars in fines and at least ten years in prison. Butler was told to pay more than \$50,000 in fines and restitution. He has already resigned from Texas Tech and paid the university \$250,000. Butler has also been stripped of his Texas medical licence. His lawyers plan to appeal.

Scientists have watched the Butler case unfold with dismay. Many believe that the federal government was harsh on Butler to create a public image that it is tough against terrorists. "Butler is going to federal prison, Texas Tech's reputation is in tatters, and no bioterrorism was even committed," says Nobel laureate Peter Agre, a former student of Butler's who is now a biochemist at the Johns Hopkins School of Medicine in Baltimore, Maryland.

"Everyone has lost," say Agre.

Some researchers say that by making an example out of a 62-year-old, respected researcher the government has undermined its own cause. "This is guaranteed to have a chilling effect on pathogen research," says Steven Block, a biophysicist at Stanford University in California. He cites examples such as that of Harvard University biochemist John Collier, who has said that he threw out his old stocks of anthrax bacteria for fear of running foul of new bioterrorism regulations.

Other researchers have denied any such problem. They point to the eager competition among many scientists for the billions of dollars in biodefence spending that have been made available to counter the threat of terrorism. ■



School for scandal: the University of California, Los Angeles, is no longer accepting body bequests.

Two arrested for trade in body parts from donor programme

Rex Dalton, San Diego

Allegations of an illicit trade in body parts at the University of California, Los Angeles (UCLA), have forced the institution's authorities to rethink the systems used to collect human tissue for research.

UCLA's David Geffen School of Medicine last week suspended its programme for accepting donated human bodies, after the arrests of an employee and an independent tissue broker. The pair are alleged to have participated in a scheme in which human body parts were supplied to biotechnology firms and research institutions.

The Willard Body Program, which receives annually about 175 cadavers for medical research and education, is a significant source of tissue for UCLA researchers, says urologist Thomas Rosenthal, associate vice-chancellor of the medical school. University officials say they are working out how to compensate for reduced access to human specimens.

Henry Reid, director of the programme, was arrested on 6 March on suspicion of grand theft. He is alleged to have allowed tissue broker Ernest Nelson to remove parts of nearly 500 cadavers from the university over the past six years. Invoices supplied by Reid show that he charged Nelson more than \$700,000 for the body parts.

Rosenthal says a preliminary investigation indicates that Reid was not paying the money

from Nelson into UCLA bank accounts. Reid, who coincidentally was hired by UCLA in 1997 to correct problems in handling cadavers, could not be reached for comment.

Nelson, who runs Empire Anatomical Services from a storage facility in a suburb of Los Angeles, shipped specimens to what his attorney, Greg Hafif, called "accredited research institutions" and US biotechnology firms. Hafif declined to name the companies, but Mitek Worldwide, a surgical-devices company based in Norwood, Massachusetts, has acknowledged receiving tissue from Nelson. Hafif says that Nelson is innocent of wrong-doing, as he believed he was working legitimately with UCLA.

Biotechnology companies and academic institutes need body parts for medical research and for training medical students. Although it is illegal to sell human body parts in the United States, firms can make a profit by charging to cover the costs of supplying specimens. A single human body can be parcelled out for thousands of dollars.

The UCLA irregularities come after similar instances in past decades at University of California medical facilities in San Diego and Irvine. Attempts over the years to enact stronger regulatory controls at the state and federal level have been stymied by a variety of political forces, including lobbying by institutions that want ready access to tissue. ■

Outspoken nuclear scientist 'forced out' over polygraph row

Jonathan Knight, San Francisco

A national security expert says he was forced to resign last year because of his vocal opposition to the use of lie detectors at his nuclear weapons lab.

Alan Zelicoff, formerly a senior scientist at Sandia National Laboratories in New Mexico, last week spoke for the first time about his resignation last July. He quit following disciplinary action against him for public criticism of polygraph testing. The lab denies any link between Zelicoff's departure and his public statements.

The Department of Energy instituted routine polygraph screening for employees at all nuclear weapons laboratories in 1999 to check for leaks of classified information. Zelicoff headed the many scientists at Sandia who objected to the move.

He published opinion pieces and cited research that suggested polygraph tests might finger innocent employees rather than catching spies.

A series of disciplinary actions followed the outbursts, culminating in a week-long suspension from the lab in June 2003, according to internal memos provided by Zelicoff. When he returned, he was barred from working on the

disease-surveillance software he had developed (see *Nature* 411, 228; 2001). The lab claimed there was a commercial conflict of interest, which he denies.

John German, a spokesman for Sandia, said that while he could not discuss Zelicoff's case, lab employees must follow procedures before speaking publicly on matters of national security. The lab memos refer to failures in this area, as well as to an alleged security infraction.

Zelicoff insists that he followed necessary procedures and says the charges in the memos are groundless. He spoke out about the affair after becoming frustrated that members of Congress to whom he had complained failed to act. Steven Aftergood, director of the Project on Government Secrecy at the Federation of American Scientists in Washington, is dismayed by Zelicoff's claims: "It tells scientists not to rock the boat." ■



Alan Zelicoff: spoke out against lie tests.

Varmus advises 'vow of chastity' over NIH staff consultancies

Meredith Wadman, Washington

Senior staff at the National Institutes of Health (NIH) should be banned from most kinds of external consultancy work, according to the organization's former director Harold Varmus.

The suggestion, which would reverse rules that Varmus himself put in place in 1995, was made on 12 March to a blue-ribbon panel set up to examine conflict-of-interest policy at the institutes. The panel was established after the *Los Angeles Times* suggested last December that some top NIH staff made biased decisions after receiving consultancy fees from drugs firms (see *Nature* 426, 741; 2003).

Varmus advised that senior staff take the same "vow of chastity" that he has taken in his current position as president of the Memorial Sloan-Kettering Cancer Center in New York. "Institute directors should probably not be consulting for any companies," says Varmus. "And certainly not for companies that might be candidates for grants from the NIH." He also included institute deputy directors, scientific directors and clinical directors in the constrained group.

Varmus acknowledged that his recommendations would severely limit the outside activities of the most senior officials. But he said other staff, including most researchers, should be exempt in order to allow the NIH to recruit and retain the best scientists. The situation should automatically be reviewed if a researcher's outside earnings reached the same levels as his or her salary, he suggested.

Varmus loosened the rules in 1995, allowing institute directors to accept consulting payments from pharmaceutical firms, and removing a \$25,000 cap on outside earnings for all NIH employees. He also allowed employees to accept outside offers of stock and stock options, a payment seen as particularly prone to tempt scientists to bias results in favour of the companies employing them. Varmus believes times have now changed. "The interface between academia and industry has increased," he says.

The panel, which is chaired by Norman Augustine, chairman of engineering firm Lockheed Martin based in Bethesda, Maryland, and Bruce Alberts, president of the National Academy of Sciences, is due to report in May. ■



Fat chance: critics say NIH director Tommy Thompson's anti-obesity campaign is misdirected.

Health experts find obesity measures too lightweight

Declan Butler

Researchers have slammed the US government's plans to combat the country's obesity epidemic. Experts say the focus on changes in personal choices is flawed, and ignores the role of government and the food industry in shaping people's decisions.

The strategy, announced on 9 March by Tommy Thompson, head of the Department of Health and Human Services, combines a nationwide education and advertising campaign with a revamp of obesity research across the National Institutes of Health (NIH).

A public-awareness campaign, dubbed 'small steps', will try to educate Americans by highlighting healthier lifestyle options, such as walking up stairs instead of taking a lift. Thompson also wants voluntary measures to encourage restaurants to put calorie information on menus and to improve the accuracy of calorie counts on food labels.

Obesity experts say that education is fine, but they are angry that the government has not addressed what they believe to be the main cause of obesity — the easy availability of cheap, calorie-rich foods. For example, a study released last month by the Center for Science in the Public Interest, a Washington-based group that campaigns on nutrition issues, showed that the main courses of children's meals in US chain restaurants typically contain 700–900 calories, more than half the total recommended daily amount.

And according to a study by the Centers for Disease Control and Prevention in Atlanta, Georgia, around 64% of US citizens are overweight or obese (A. H. Mokdad *et al.* *J. Am. Med. Assoc.* 291, 1238–1245; 2004), with obesity accounting for 400,000 preventable

deaths in 2000, second only to tobacco.

Regulating the food industry is the best way to tackle the problem, say public-health experts and nutrition scientists. They recommend banning soft drinks and fast foods in schools and food advertising directed at children. They also want compulsory labelling of the calorie content of restaurant menus.

"The campaign says nothing about what the government or the food industry could do to help people to eat less and move more," says Marion Nestle, chair of the Department of Nutrition, Food Studies and Public Health at New York University.

The other component of Thompson's plan — changes to NIH obesity research — has been welcomed by obesity experts. Under the scheme, put together by a cross-agency task force set up by NIH director Elias Zerhouni, the institutes will focus on behavioural and environmental approaches to modifying lifestyle. It will also support clinical trials of diet strategies (see page 252).

Increased focus on research to prevent obesity is essential, says Kelly Brownell, director of Yale University's Center for Eating and Weight Disorders. He points out that the NIH has been criticized for focusing too much on the genetics, pharmacology and treatment of obesity. Research in these areas will continue under the plan, which could be funded by a \$40-million increase for obesity included in the 2005 NIH budget request.

But many scientists worry that marking obesity as a research problem risks 'medicalizing' the issue and reinforcing public belief that it is a disease requiring treatment. ■

▶ <http://cspinet.org/new/pdf/kidsrestfood.pdf>

▶ <http://obesityresearch.nih.gov>

Pentagon attempts to bend light to its will

Geoff Brumfiel, Anaheim

A research initiative is being launched by the Pentagon to investigate negatively refractive materials — structures that seem to defy basic physics by bending light the ‘wrong’ way.

A chunk of negatively refracting material about the size of a cheese wedge was on display at a technical meeting held last week by the Defense Advanced Research Projects Agency (DARPA) in Anaheim, California. DARPA officials say that it could be used in a variety of military technologies, including radar, communications and optics for lasers and imaging.

“We’re looking to identify applications,” says Valerie Browning, a programme manager at DARPA’s Defense Sciences Office in Arlington, Virginia. The agency will unveil plans to study and apply these materials in military systems in the next month, Browning says.

Negative refraction was conceived in the 1960s, when theorists claimed that some materials should bend light in a strange way. When a beam of light passes from one material, such as air, to another, such as glass, it bends. This refraction is how glass lenses focus light. Negative refractive materials should bend light in the opposite direction to other substances (see diagram).

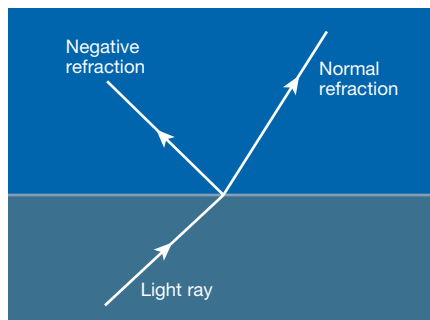
The idea remained speculative until 2001, when physicists at the University of California, San Diego, announced that they had made such a material. The structure was a ‘metamaterial’, an array of sheets of copper rings and wires, which could negatively refract microwaves.

Although some physicists were sceptical about the results (see *Nature* 420, 119–120; 2002), subsequent experiments have led most to accept the finding, says John Pendry, a physicist at Imperial College in London who works on negatively refracting materi-



Fresh angle: metamaterials such as this copper-based structure have strange optical properties.

als. For example, George Eleftheriades and Anthony Grbic, physicists at the University of Toronto, have recently created a microwave lens from a negatively refracting structure (G. Eleftheriades and A. Grbic *Phys. Rev. Lett.* in the press).



Negative refraction bends light in an unusual way.

Browning says that such advances convince her that the effect is real. But she still has questions about how the materials work. For example, some predict that the structures will amplify light waves that are normally dissipated by glass optics. If true, the material could make lenses that are more powerful than any possible with conventional optics. “But we’re still unsure of the physics,” Browning adds.

The new initiative will explore the physics and investigate the uses of negatively refracting materials in real systems. Researchers at the Boeing Corporation in Seattle are already investigating negatively refracting materials as optics for steering radar and radio communications, Browning says. And if new materials can be developed to bend infrared or even visible light, the applications could be even broader.

Biomedical institute wins reprieve from relocation

Laura Nelson, London

Plans to move Britain’s National Institute for Medical Research to a site outside London have been scrapped, according to a task force run by the institute’s funders, the Medical Research Council (MRC).

The announcement is a partial victory for the institute’s researchers, who protested last year when the MRC suggested moving the centre to Addenbrookes Hospital in Cambridge. The move was intended to improve links between clinical and basic researchers (see *Nature* 423, 573; 2003).

But it will not end the uncertainty surrounding the future of the institute,

based in Mill Hill in north London. The task force says that the centre needs to support more clinical studies, and may be better off closer to London hospitals.

The group was set up after last June’s protests, and is chaired by MRC chief executive Colin Blakemore. Results of an 8 February meeting, published on 4 March, show that the task force rejected options to close the institute, move it, or sell its site and spread the work around other London institutes.

But the group has not decided whether the institute will stay where it is, with added facilities for clinicians, or move to a more

central site. “Both options require substantial investment,” notes taskforce member Robin Lovell-Badge, a developmental geneticist at the institute. Funding issues will be discussed at the next meeting on 18 April.

The need for greater clinical integration, emphasized by task-force suggestions that the institute’s name be changed to the MRC National Institute for Human Health, has already been made a priority by Blakemore. The move mirrors the situation in the United States, where the National Institutes of Health is attempting to channel basic research in areas such as genomics into clinical advances.

Physicist accused of burning vehicles in environmental protest

San Diego A physics doctoral student at the California Institute of Technology (Caltech) in Pasadena was charged last week with arson for torching 125 vehicles in an environmental protest. William Cottrell was arrested after FBI agents traced e-mails he is alleged to have sent from Caltech's library.

Cottrell is accused of causing \$3.5 million worth of damage to vehicles, mainly sports utility vehicles, at four car dealerships in August 2003. Spray-painted lettering attributed the fires to the Earth Liberation Front, a group linked to attacks across the United States. Cottrell was denied bail at a hearing in a US district court in Los Angeles.

NASA under pressure to offer Hubble more scope

Washington The struggle by astronomers to save the Hubble Space Telescope from a premature end gained momentum last week. Senator Barbara Mikulski (Democrat, Maryland) convinced NASA to commission a detailed study of the risks of a manned mission to maintain the telescope.

NASA announced in January that it was

scrapping a planned shuttle mission to repair Hubble in 2006 — a decision that will end the telescope's working life sometime after 2007.

On 11 March, NASA asked the National Academy of Sciences to investigate this move. But later that day, the agency's administrator, Sean O'Keefe, told reporters he saw little chance of reversing his decision to scrap the repairs. This prompted Mikulski to send O'Keefe a letter warning him not to prejudge the outcome of the study. But O'Keefe stuck to his guns, arguing in the *New York Times* on 14 March that "it may not make sense to devote time and energy to a [manned] mission to the Hubble"; although robotic servicing remains a possibility.

China aims high with Everest climate station

Tokyo Those brave or foolish enough to climb to the top of Mount Everest will have better weather guidance later this year when China sets up the world's highest climate station, at an altitude of 5,200 m, on the Tibetan side of the mountain.

The unmanned station will record data on wind speeds, air pressure, rainfall, temperature and humidity, primarily for long-term climate monitoring of regional weather patterns, according to China's state media Xinhua last week. It may also be of



Looking up: a climate station on Mount Everest will give climbers better weather forecasts.

use for climbers — 175 of whom have died trying to reach the summit since 1922.

Weather in the region is currently monitored by a station 90 km away, making the unpredictable weather even more difficult to forecast. The station will be part of a 27-million-yuan (US\$3-million) network of 41 satellite-linked automatic monitoring stations that are being built across Tibet as part of an effort to develop the region.

Danish government leaves Lomborg in the clear

Copenhagen Political scientist Bjørn Lomborg no longer stands accused of bad scientific practice by publishing his book

The Skeptical Environmentalist, which questioned claims about the extent to which human activities are causing environmental damage.

Following allegations that the book sought to mislead, the Danish Committees on Scientific Dishonesty concluded in January 2003 that Lomborg had selectively quoted literature that supported his arguments, passing off his work as a formal scientific analysis (see *Nature* 421, 201; 2003). But in January 2004, the Danish government, which appointed Lomborg in 2002 as head of the Environmental Assessment Institute in Copenhagen, asked the committee to reconsider.

The committee decided last week not to repeat the investigation, saying it can only investigate allegations of scientific misconduct and had concluded that Lomborg was guilty only of bad scientific practice. In the meantime, the government had annulled the committees' original decision, leaving Lomborg's reputation officially intact.

Socialist win stems row over Spanish laws

Madrid The Socialist Party's unexpected victory in the Spanish general elections on Sunday has alleviated fears for the future of stem-cell research in Spain.

Sinking beer bubbles turn physics on its head

London Beer lovers are not suffering the effects of their tittle when they claim that air bubbles in a pint can go down, as well as up. Dick Zare, a chemist at Stanford University, California, used a high-speed digital camera to show that bubbles actually fall in a pint of the St Patrick's Day favourite, Guinness.

After a pint has been pulled, the bubbles rise more easily in the centre of the glass where there is less drag, explains Zare. This pushes beer to the top of the glass, from where it cascades back down the sides, carrying smaller bubbles with it. This alcoholic conveyor belt continues until all the bubbles settle into the creamy head that tops the Irish brew.

The same effect even happens in fizzy water, although emptying out the reaction vessel is generally considered less fun.



Before the election, scientists in Andalusia in southern Spain were concerned about a clash between regional and national law over the use of embryonic stem cells. The national law is the more restrictive of the two — for example, it states that embryonic stem cells should be used only as a last resort (see *Nature* 428, 7; 2004). A tribunal, scheduled for the summer, was to resolve whether the regional or the national law would prevail.

Now that José Luis Rodríguez Zapatero has swept to power, the Socialist Party rules at both the regional and national level. It is thought that the tribunal will be postponed, and that the Andalusian law will continue to apply.

"We are very happy," says Francisco Martin, a stem-cell researcher at the Bioengineering Institute at Miguel Hernández University in Alicante, who uses facilities in Andalusia.



Mission impossible?

Mounting criticism of the way the Bush administration handles scientific advice has put John Marburger, the US president's science adviser, in the hot seat. Geoff Brumfiel takes the temperature.

When John Marburger took over as head of Brookhaven National Laboratory in New York State six years ago, the lab was facing a public-relations disaster — a leak from a storage pool for nuclear waste had angered environmentalists and turned many local people hostile. Using a combination of patience and openness, Marburger got the facility back on an even keel.

But Brookhaven's travails were a storm in a teacup compared with the tempest the 63-year-old physicist now faces as scientific adviser to a Republican president. On 18 February, the Union of Concerned Scientists (UCS) issued an impassioned statement signed by more than 60 eminent researchers, accusing the Bush administration of misrepresenting scientific findings, suppressing reports and loading scientific advisory panels with its political supporters. "The pervasive, systematic nature of these practices is unprecedented," says Kurt Gottfried, a physicist at Cornell University and chairman of the Cambridge, Massachusetts-based group.

The statement brought to a head a range of criticisms of President George W. Bush

that have been simmering in the scientific community since he came to office in January 2001. His administration has withdrawn from the Kyoto Protocol on climate change, which was supported by environmental scientists. It has also pressed ahead with missile-defence systems that top US physicists said won't work, and set rules for stem-cell research that have riled leading biologists. In addition, it has selected members for scientific panels whom many researchers see as ideologues. Marburger, and the White House's Office of Science and Technology Policy (OSTP) which he directs, must now defend these decisions, and that charge has fuelled a debate about the role that he and the office play in the current administration.

On the outside

The OSTP is a relatively small outfit of about 60 specialist staff whose job it is to prepare advice for the president and coordinate science policy across the federal government. It was thrust on an unwilling White House by Congress in 1976 and has struggled ever since to exercise real clout. Huge government departments, such as

health and defence, direct most of the on-the-ground research, whereas the setting of science-related policies on sensitive topics, such as global warming and stem-cell research, rests chiefly with political figures in the administration. "We were the bastards at the family reunion in many ways," recalls John Gibbons, who ran the office from 1993 to 1998 for President Bill Clinton. "We were always on the outside."

From the beginning, Marburger seemed more on the outside than most. He arrived in Washington ten months after Bush took office and lacked the title of 'special assistant' to the president, which was given to previous science advisers. As a registered Democrat with few contacts in the administration, the fear was that Marburger would have little access to the president — the milestone by which many judge the success of the science adviser.

But some science advocates saw cause for optimism. Historically, the science adviser has been stuck between scientists, who want someone to represent their interests in the corridors of power, and the president, who wants someone to defend policies and deflect researchers' habitual moaning about money.



Prominent US researchers claim that John Marburger (left) lacks influence at the White House.

Under Republican administrations, the challenge of meeting both sides' expectations is particularly acute as scientists, for the most part, lean to the political left. Marburger, whom colleagues invariably describe as friendly and approachable, seemed in many ways to be the perfect person for the job. He was seen as a consensus builder, responsive to the concerns of the outside, but willing to work within the rules laid down by the president.

An open ear

Two-and-a-half years later, the verdict on his performance is mixed. In issues ranging from visas for foreign scientists to restrictions on publishing sensitive information, advocacy groups have found a receptive ear in Marburger, says Janet Shoemaker, head of public affairs at the American Society for Microbiology. But many doubt whether he has the clout in the administration to act on the input he receives. "It's been clear from the beginning that the office was not going to be an essential part of White House operations," says Mike Lubell, who heads public affairs at the American Physical Society. "Science just isn't high on the president's personal agenda."

Marburger bristles at the suggestion that his office lacks influence inside the White House. "I sit in on all the senior staff meetings every day at 7.30," he says. Although his office may eschew high-profile initiatives, it has been effective in improving coordination between agencies and securing more funding for science, he argues. Marburger says that what he lacks in Washington ties, he makes up for with his management skills, which he acquired during 14 years as president of the State University of New York in

Stony Brook and six years as director of Brookhaven. "This president came in with a business point of view, and he hired people sympathetic to that point of view," he says. "I don't think any previous science adviser, in the past two decades anyway, has had as much executive experience as I have."

Former OSTP staff and administration officials outside his office confirm that Marburger has worked well with others in the government. The office has been particularly active in defining the science and technology needs of the Bush administration's top priority — the war on terrorism. One such example came earlier this month, when Marburger helped to outline an approach for the funding and publication of biological research that might be abused by terrorists (see *Nature* **428**, 109; 2004). The result was a flexible plan that biologists have endorsed, not the draconian one some of them had feared from the security-conscious administration.

Marburger has also played a strong role in setting the budget, according to science lobbyists and congressional staff. Officials at the White House Office of Management and Budget — which writes the president's annual budget proposal — say that Marburger has been influential in budget negotiations, even winning increases for some agencies in the final stages of budget wrangling. "Jack may not get everything he wants, but he always gets something," a senior official there says.

But on political issues, such as the UCS report, Marburger seems less adept. The report was largely a rehash of complaints that had been building throughout the administration, yet Marburger seemed blind-sided by its publication. In a hastily convened press conference afterwards, he called the group's

claims a "conspiracy theory" and "a very deliberate attempt to undermine the president in a campaign year", but so far he has not provided a rebuttal to each of the statement's claims. A detailed response is expected soon.

His aggressive, yet unspecific response has drawn criticism of his performance, even from Republicans. "The way they have handled this has not reflected well on the office," says one congressional staffer. Senator John McCain (Republican, Arizona), chair of the Senate Committee on Commerce, Science and Transportation, and perhaps Bush's most persistent critic in his own party, has scheduled hearings on the UCS report. And Senator Christopher Bond (Republican, Missouri) told Marburger at a 26 February budget hearing that he was "very concerned about these accusations".

Fighting back

Marburger continues staunchly to defend the administration's approach to scientific issues. "We're not trying to make things consistent with some ideology," he says. He also complains that some problems could have been brought to his attention earlier. Referring to one of the UCS statement's allegations, that the National Cancer Institute posted outdated studies linking abortion to breast cancer on its website, he says: "Do dumb things happen from time-to-time? Yes. But if people call me up, I do something about it. We respond to evidence of problems." But the impression that the OSTP lacks influence has made outsiders reluctant to bother it with their complaints. "I don't think they have the power to intervene," says Gottfried.

The issues raised in the UCS statement look set to dog Marburger's office until November's presidential election. Earlier this month, the removal of Elizabeth Blackburn, a biologist at the University of California, San Francisco, from Bush's Council on Bioethics alarmed many scientists (see *Nature* **428**, 4; 2004), and Gottfried says that so far, more than 1,000 researchers have added their signatures to the UCS statement.

It's all a far cry from Brookhaven, where Marburger's unassuming, low-key style and his willingness to work with local environmentalists quickly drew the sting from the lab's most strident critics. In Washington, the political game is much tougher, and some senior figures are prepared to discuss the possibility that Marburger might resign. "I don't believe I would have been able to stay in that office under a presidency that has done some of the things that have been done," says Gibbons. Yet many of the administration's harshest critics do not blame Marburger personally. "Our purpose is not to make his life more difficult," says Gottfried. "There are deeper forces at work here, and science is only part of it." ■

Geoff Brumfiel is *Nature's* Washington physical sciences correspondent.

♦ www.ostp.gov

Slim pickings

The dieting industry is a massive money-spinner. Yet across the developed world, waistlines continue to expand. Declan Butler examines the sparse scientific evidence behind the claims made for leading diet plans.

“All my friends are on the Atkins diet, and they’re losing weight like crazy. Would this be a good thing for me or not?” Four years ago, Dena Bravata, a physician at Stanford University in California, could only stare blankly at her querying patient. She was aware of the popular low-carbohydrate diet, but she didn’t have a clue what to suggest. “That’s what motivated me to do a study,” Bravata says. “I had lots of patients all asking me the same question.”

So Bravata called her twin sister Dawn at Yale University in New Haven, Connecticut. After roping in further colleagues, they began scouring the scientific literature as far back as 1966 for papers addressing the efficacy and safety of low-carbohydrate diets. In April last year, the researchers revealed the results of this systematic review¹, which should provide those embracing the Atkins craze with food for thought. They couldn’t find any evidence that diets containing low proportions of carbohydrates were more effective. Consuming fewer calories overall, and sticking with a diet for a long time, were the only significant factors for slimming success.

The survey also revealed how painfully thin the science of dieting is. Searching for relevant terms on literature databases turned up a total of 2,609 articles. But just 94 of the studies met the researchers’ criteria for inclusion in their review, such as using proper controls and lasting longer than four days. In few cases were there sufficient data to assess the safety of the diets. Information on exercise was rarely reported, making it impossible to compare diets in sedentary and active patients. Few studies looked at different ethnic groups, only five lasted longer than 90



Food fight: do commercial diet plans have sufficient scientific nous to beat the bulge?

days, and not many included participants over 60 years old. If you are planning to diet in later life, there is little evidence to go on.

Given the burgeoning obesity epidemic, that’s a sorry state of affairs. Two-thirds of US adults are overweight, a third are clinically obese, and 1 in 12 has diabetes — a common complication of weight problems. The incidence of obesity in the United States has doubled in the past decade alone, according to the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. And where the United States has led, the rest of the world is following.

Eat yourself fitter

Hundreds of diet books claim that the answer lies with their particular prescription for tweaking the proportions of fat, carbohydrate and protein that we eat. Some focus on the idea that certain forms of carbohydrate — those with a low ‘glycaemic index’, which are metabolized more slowly — are more conducive to weight loss than others (see ‘Good carbs, bad carbs’, overleaf).

A multibillion-dollar industry hangs on these claims. Robert Atkins, the New York doctor who championed the low-carbohydrate approach, is no longer with us, having perished last April after slipping on an icy street. But Atkins Nutritionals, the

company he founded to market his diet and related products, is very much alive — in October, two leading investment firms together paid a sum rumoured to be as high as \$800 million for a controlling stake.

But do any of these diets work? And are they safe? “The public is frantic for a sane voice amid the cacophony of popular diets,” says Marion Nestle, who chairs the Department of Nutrition, Food Studies and Public Health at New York University.

Science has so far had little to say. Only recently have research agencies shown any interest in funding trials of recipes for weight loss. Many of the studies reviewed by Bravata and her colleagues were initiated to examine the effects of varying the diet on particular diseases. What we need, say experts on nutrition, are much bigger trials that are specially designed to address the issue of weight loss and that meet the standards of those used to test new drugs.

One month after Bravata’s review appeared, two randomized controlled trials of low-carbohydrate diets were published back-to-back in *The New England Journal of Medicine*. One was led by Frederick Samaha at the Veterans Affairs Medical Center in Philadelphia², the other by Gary Foster at the University of Pennsylvania in the same city³.

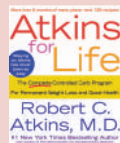
Samaha and Foster each randomly

Something to chew on

Diet plan

Atkins diet

♦ atkins.com



Sample claim

"...your body converts from the metabolic pathway of burning carbohydrate to burning fat as the primary energy source. This results in weight loss."

Eat less

Rice, whole grains, vegetables, dairy products

Eat more

Beef, poultry, soya protein

South Beach diet

♦ www.southbeachdiet.com

"[The South Beach diet] teaches you to rely on the right carbs and the right fats... [It] will not only help you lose weight but also improve your health."

Fatty meats, potatoes, white bread, soft drinks, refined grains, sweets

Poultry, lean meat, whole grains, fish, fruit and vegetables

New Glucose Revolution

♦ www.glycemicindex.com



"[Low-glycaemic-index diets] have benefits for weight control because they help control appetite and delay hunger."

Potatoes, white bread, soft drinks, refined grains, sweets

Breakfast cereals, wholemeal bread, fruit and vegetables

Weight Watchers

♦ www.weightwatchers.co.uk

"You are assigned a daily food Points allowance. Keep within your Points allowance and lose weight. It's that simple!"

Reduced intake of most food groups

Fruit and vegetables, poultry, seafood, grains, lean meats

treatment. "We have moved it from quackery to science," says Foster. "Maybe there is something there, with the emphasis on the 'maybe'."

The trials also failed to find signs that cutting carbohydrates rather than fats will increase the risk of cardiovascular disease. If anything, levels of artery-clogging triglycerides, and of 'good' versus 'bad' cholesterol, were slightly better among the volunteers on the low-carb diets.

But the studies were small: Samaha's included 132 patients, Foster's just 63 — and only 79 and 37, respectively, saw the studies through to the end. Until longer-term data from larger trials are in, Foster is loath to recommend any low-carbohydrate diet. He is now starting a larger study in 360 patients, funded by the US National Institutes of Health, which will last for two years. "Our initial study was just a pilot project; this is a full-blown clinical trial," says Foster. It will not only address the big question of whether low-carb or low-fat diets are better at maintaining weight loss in the long term, but will also include a wider assessment of the volunteers' health. Foster will collect data on the patients' ability to exercise, and on the health of their arteries, bones and kidneys.

This is important, because many nutritionists remain concerned about the safety of long-term adherence to diets that are biased heavily towards fats and proteins. One recent review of the safety of low-carbohydrate diets⁴ reeled off an alarming list of potential problems: "Complications such as heart arrhythmias, cardiac contractile function impairment, sudden death, osteoporosis, kidney damage, increased cancer risk, impairment of physical activity and lipid abnormalities can all be linked to long-term restriction of carbohydrates in the diet."

Unbalanced diet?

Robert Eckel of the University of Colorado's Center for Human Nutrition in Denver is concerned that people following the Atkins diet, or similar plans, won't eat the wide range of fruits, vegetables and whole grains that is almost universally held to be beneficial. "The Atkins diet is anything but health-promoting," he argues.

Other experts argue that the debate over the relative merits of various diets is a sideshow to the main message for those wanting to lose weight and keep it off: consume fewer calories and exercise more. "Thinking that a specific diet should eliminate people's weight problems is totally unrealistic," asserts Arne Astrup, a nutritionist at the Royal Veterinary and Agricultural University in

assigned obese volunteers to either an Atkins-style low-carbohydrate or a traditional low-fat diet. In Samaha's six-month study, subjects on the low-carb diets lost the most weight — an average of about 6 kilograms, roughly three times the weight loss reported for the low-fat group. Foster's smaller, year-long trial initially showed similar results, but the difference between the two groups had disappeared by the end of the year.

These two studies provide the best evidence so far that low-carbohydrate diets may be of some use for patients who need to shed weight — at least in the early stages of their

Frederiksberg, Denmark, and president-elect of the International Association for the Study of Obesity. "There is no getting round the laws of thermodynamics." In other words, if your energy intake exceeds your energy output, you will get fat whatever the proportions of fat, carbohydrate and protein in your diet.

The svelte Astrup practises what he preaches, cycling a total of 25 kilometres to and from work each day. "That burns up about 800 calories," he says. "As a result, I can enjoy eating without ever thinking about it."

Hard to swallow

Studies on people in hospital wards where their food intake is strictly controlled reinforce Astrup's message. These show that weight loss is similar for diets containing the same number of calories, irrespective of the proportion of carbohydrate, fat and protein^{5,6}.

But some experts, including researchers who are otherwise critical of the Atkins diet, say that the low-carb approach at least has the merit of encouraging people to cut out large portions of calorie-rich foods such as doughnuts, cakes and pastries. That may result in dieters consuming fewer calories overall, which would explain the results of Samaha's and Foster's trials.

But it is difficult to determine how many calories patients in dieting trials are consuming. Many volunteers don't stick to their prescribed diets, and they typically under-report the amount that they have eaten.

"The main reason the science is so poor is the same reason why we study the issue in the first place — most people don't like restricting their food intake," says Sandi Pirozzo, an epidemiologist at the University of Queensland in Brisbane, Australia. Pirozzo is the lead author of a recent review suggesting that the low-fat diets that have been recommended by doctors for decades seem to be no more successful than alternative diet plans⁷.

Another way to examine the value of various approaches to weight loss is to start with success stories — people who have lost large amounts of weight and kept it off — and look for common factors. This approach can be useful for generating hypotheses to test in subsequent controlled clinical trials, and has been used by Rena Wing, a psychiatrist at the Lifespan Academic Medical Center at Brown University in Providence, Rhode Island, and James Hill of the University of Colorado's Center for Human Nutrition.

A taste of success

Wing and Hill's National Weight Control Registry tracks the habits of nearly 3,000 successful dieters. People who can manage their weight, Wing and Hill suggest, share four common factors: they are on low-fat diets, closely monitor their weight and food consumption, exercise for more than one hour a day, and don't skip breakfast⁸. These

Good carbs, bad carbs

One of the hottest debates in dieting science is whether carbohydrates that are slowly broken down into glucose are better for people wanting to lose weight than those that are metabolized more rapidly.

Carbohydrates can be rated according to their glycaemic index (GI), a measure proposed in 1981 by David Jenkins of the University of Toronto in Canada¹⁰. This is based on the rate of digestion of a given carbohydrate and its conversion into blood glucose.

High-GI foods, such as white bread and potatoes, provoke a rapid post-meal spike in blood glucose and insulin, which subsequently converts blood sugar into long-term energy stores. Foods with a low GI, such as lentils, breakfast cereals and brown rice, cause a much lower and smoother response. Diabetics are often advised to eat low-GI foods, and there is evidence that these carry lower risks for conditions such as heart disease.

But can low-GI diets help to reduce weight? Some studies show that low-GI foods seem to make people feel less hungry, an effect that is possibly linked to the smoother rise and fall of blood insulin levels^{11,12}. The spike in insulin triggered by high-GI foods might also direct glucose away from being burned in muscle towards being stored as fat.

But most studies have been based on the consumption of simple foods, rather than the complex mixed meals that we eat in the real world. This is where the picture gets really messy. The total glycaemic load for a particular



meal can depend not just on the carbohydrates present, but also on their interaction with fat and proteins, and even how the meal is prepared¹³.

Sceptics of the relevance of GI to dieting, such as Arne Astrup of the Royal Veterinary and Agricultural University in Frederiksberg, Denmark, argue that it may be impossible to predict the glycaemic load of a mixed meal with sufficient accuracy to be of practical use. But David Ludwig, an endocrinologist at Harvard Medical School in Boston, disputes this view. He believes that the best approach for weight loss is a diet that stresses low-GI carbohydrates and contains moderate amounts of fat — a compromise between traditional low-fat diets and the Atkins low-carb approach.

The jury is still out, pending the results of clinical studies. Ludwig has reported encouraging results in a pilot project involving 16 adolescents¹⁴ and is now carrying out an 18-month trial on 100 obese children, randomly assigned to either a low-GI or a conventional low-fat diet.

findings suggest that behavioural factors may be at least as important in determining weight loss as the proportions of fat, carbohydrate and protein in the diet.

Nutritionists agree that the rising tide of obesity won't be reversed without concerted action by governments and the food industry to promote the consumption of foods that are less energy intensive. Last month, the CDC estimated that between 1971 and 2000, the calorie intake of US men rose 7%, and that of women jumped 22% (ref. 9). Most of this increase can be attributed to a boom in eating out, snacking, the consumption of soft drinks and bigger portion sizes. Twenty years ago, a typical US bagel was 3 inches across and contained 140 calories; now it has twice the diameter and packs a whopping 350 calories.

But if doctors are going to be able to offer sound advice to those who have already become dangerously overweight, more research into dieting will be vital. Foster argues that the public and the media need to change their conception of the problem. Rather than simply asking whether diet A is generally 'better' than diet B, he suggests, we should move towards a situation in which doctors can tailor diets to individuals, depending on their metabolic and behav-

ioural characteristics. Different diets might also have to be used at different stages in a patient's treatment.

"Ideally, five years down the road, with lots of scientific data behind us, we would be in a position to recommend diets tailored to individuals such as, say, diabetics, meat lovers, or those with difficulty sticking to a low-fat diet," says Foster. "This whole idea of one diet wins, one diet loses, seems to say that every overweight person is the same behaviourally and metabolically, and that is just silly."

Declan Butler is Nature's European correspondent.

1. Bravata, D. M. *et al.* *J. Am. Med. Assoc.* **289**, 1837–1850 (2003).
2. Samaha, F. F. *et al.* *N. Engl. J. Med.* **348**, 2074–2081 (2003).
3. Foster, G. D. *et al.* *N. Engl. J. Med.* **348**, 2082–2090 (2003).
4. Bilsborough, S. A. & Crowe, T. C. *Asia Pacific J. Clin. Nutr.* **12**, 396–404 (2003).
5. Golay, A. *et al.* *Am. J. Clin. Nutr.* **63**, 174–178 (1996).
6. Golay, A. *et al.* *Int. J. Obes. Relat. Metab. Disord.* **20**, 1067–1072 (1996).
7. Pirozzo, S., Summerbell, C., Cameron, C. & Glasziou, P. *Obes. Rev.* **4**, 83–90 (2003).
8. Klem, M. L., Wing, R. R., McGuire, M. T., Seagle, H. M. & Hill, J. O. *Am. J. Clin. Nutr.* **66**, 239–246 (1997).
9. *MMWR Morb. Mortal. Wkly Rep.* **53**, 80–82 (2004).
10. Jenkins, D. J. *et al.* *Am. J. Clin. Nutr.* **34**, 362–366 (1981).
11. Pawlak, D. B., Ebbeling, C. B. & Ludwig, D. S. *Obes. Rev.* **3**, 235–243 (2002).
12. Raben, A. *Obes. Rev.* **3**, 245–256 (2002).
13. Pi-Sunyer, F. X. *Am. J. Clin. Nutr.* **76**, (Suppl.) 290S–298S (2002).
14. Ebbeling, C. B., Leidig, M. M., Sinclair, K. B., Hangen, J. P. & Ludwig, D. S. *Arch. Pediatr. Adolesc. Med.* **157**, 773–779 (2003).

Concern is more than just 'ruffled feathers'

If a government abuses science to justify its policies, scientists have a duty to speak out.

Sir—Your News story “Scientists slam Bush record” (*Nature* 427, 663; 2004) reports on the statement by 63 prominent scientists accusing the Bush administration of “misrepresenting and suppressing scientific knowledge”. John Marburger, the administration’s head of science and technology policy, quickly responded to the initiative from the Union of Concerned Scientists (UCS) by dismissing the move as political and simply the result of a few individuals having their “feathers ruffled”, according to the *New York Times*.

A similar response greeted Congressman Henry Waxman’s like-minded report last August.

The administration’s response to the UCS initiative shows that nothing

short of a broad-based condemnation will deter this administration’s misuse of science.

We are PhD students and postdocs at Stanford and the University of California, Berkeley, who are attempting to publicize the widespread alarm of scientists at the Bush administration’s use of science (www.scienceinpolicy.org). We have examined a broad range of environmental issues and uncovered a pervasive pattern of misuse, suppression and contradiction of science, including that performed by the administration’s own researchers.

This is not about a few “ruffled feathers”. At the time of writing, more than 1,000 scientists, from all 50 US

states and from around the world, have signed our statement decrying the Bush administration’s misuse of science.

Many of us are publicly funded researchers who feel that, if the current US administration is abusing science to justify its policies, we have a moral responsibility to speak out.

We invite you to join in these efforts to restore scientific integrity to US policy-making.

Kai M. A. Chan*, **Stephen Porder***,
Paul A. T. Higgins†, **Sasha B. Kramer***

*Department of Biological Sciences,
Stanford University, Stanford,
California 94305-5020, USA

†151 Hilgard Hall, University of California,
Berkeley, California 94720-3110, USA

Health-aid efforts rely on local infrastructure

Sir—Reading your News story “The fightback starts here” (*Nature* 426, 754; 2003), it is gratifying to see that major organizations and donors are focusing on the health problems of developing countries. Perhaps improved interventions will address diseases such as AIDS, tuberculosis and malaria, but it is not always the lack of technologies and drugs that constitute the problem. As one who has been involved in this battle for many years, I know that improvements in basic public health are also needed.

To maximize the effect of the Global Fund and resources from other donors, a matching international and African effort is needed to refurbish basic health systems. For example, some 50 years ago malaria and tuberculosis in particular were under fairly effective control, at least outside sub-Saharan Africa. Even now, with techniques already on hand, an integrated approach involving indoor application of residual insecticides and using diagnosis and effective treatment would bring malaria under control in many countries. Proper planning and better public health is also needed to improve defective services, such as water and sanitation, which contribute so greatly to developing-world problems.

I hope that international donors will encourage the governments of afflicted countries to take specific steps: to adopt public-health principles by action rather than words; to foster and sustain local expertise in research and development; to show commitment to the well-being of the public and accept that sustainable health

programmes require long-term government support.

The issues are both political and financial, and they cannot be ignored indefinitely. Effective public-health infrastructures are essential to the health of the communities that the Global Fund and other agencies wish to assist. These donors could use their influence to ensure that governments develop and use effective health planning. Their objectives should be the rebuilding of local health infrastructures that are sustainably funded, functional and transparent in decision-making. With such changes in hand, these international efforts might have a real chance of succeeding.

Clive Shiff

Johns Hopkins Malaria Research Institute,
615 North Wolfe Street, Baltimore,
Maryland 21205, USA

Confidential reports may improve peer review

Sir—Jean-Patrick Connerade in Correspondence (“Scandals stem from low priority of peer review” *Nature* 427, 196; 2004) offers an excellent perspective on the current status of the peer-review system.

Although I agree with his main argument that refereeing needs further recognition, I disagree with the suggestion that recent cases of misconduct, plagiarism and other problems arise from the low priority of peer review.

Most referees are indeed very busy, but there are good and competent scientists who still provide a critical,

detailed and timely analysis. Last-minute and ill-conceived reports are unjustified, because there is no obligation to act as a referee. In their letters to reviewers, editors always point out that the potential referee should decline if they are unable to reply in a timely fashion.

It’s true that peer review is time-consuming, but information provided by editorial offices about individual reviewers could become an additional and useful measure of a scientist’s professional qualifications.

Some journals already do this. For example, the Royal Society of Chemistry has for many years provided annual confidential reports to its referees that list the papers refereed for each journal, the referee’s recommendation in every case, and the final outcome. Not only does this provide useful feedback for the reviewers, it can help to normalize reviewing standards and even improve the quality and speed of assessments. The Royal Society of Chemistry is currently updating its system so that reviewers will be able to access this confidential information online.

As more and more journals develop electronic systems of publication and record-keeping, this sort of information should be easy for them to provide.

Pedro Cintas

Department of Organic Chemistry,
Faculty of Sciences, Universidad de Extremadura,
E-06071 Badajoz, Spain

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

Shooting time's arrow

The laws of physics mean that the Universe is slowly dying.

The Fabric of the Cosmos: Space, Time and the Texture of Reality

by Brian Greene

Alfred Knopf: 2004. 512 pp. \$28.95

Allen Lane: 2004. £25

Paul Davies

The most apocalyptic pronouncement in the history of science was made by the physicist Hermann von Helmholtz in 1854. The Universe, von Helmholtz declared, is dying. His pessimistic prediction was based on the second law of thermodynamics, according to which the entire cosmos is on a one-way slide towards a state of complete disorder, or maximum entropy, from which it will be unable to extricate itself.

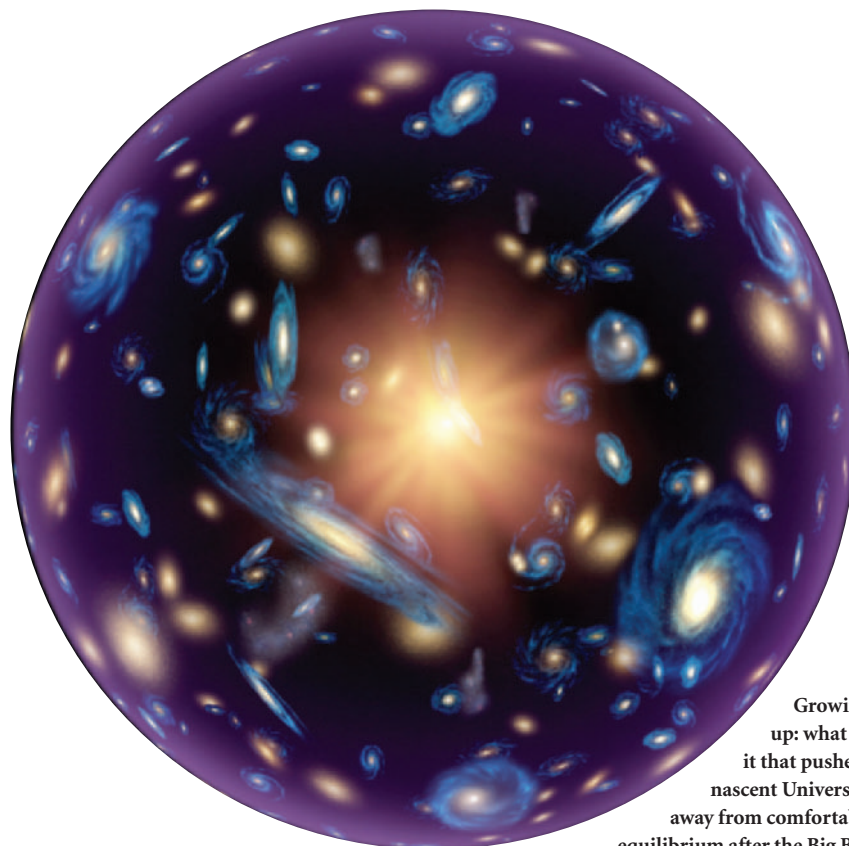
The basis of the dying Universe is easy to explain. We are surrounded by processes that have a definite directionality to them: people grow old, cars rust, cliffs are eroded, sandcastles get washed away. Take a movie of any familiar scene and run it backwards — people will laugh, because the world in reverse looks so preposterous.

The same directionality pervades the cosmos. The Sun, for example, is not the immutable orb it may seem. Over billions of years it has consumed much of its fuel, and eventually it will burn out and die. The same goes for all stars. Irreversible decay and degeneration spell universal heat death, just as von Helmholtz proclaimed.

But this incontestable fact conceals a deep mystery. The world about us may display a conspicuous 'arrow of time', but the underlying laws of physics are, with a minor exception, completely symmetric. At the level of individual atoms, there is nothing to distinguish future from past. So somehow, between atom and cosmos, the temporal lopsidedness of the Universe emerges. How this happens is one of the longest-running debates in science.

I became fascinated by this mystery after hearing a lecture delivered at the Royal Society in London by Fred Hoyle in 1968, and I have returned to it again and again throughout my career. Early on it became obvious to me that the subject was marred by widespread misconceptions and muddled thinking, even among some of the giants of theoretical physics. It is therefore extremely heartening that Brian Greene has written such a sensible and consistent account of this tantalizing topic, where so many other authors have merely added to the confusion.

Following the success of his book on the unification of physics, *The Elegant Universe*, Greene has turned his attention to the nature of space and time. Before zeroing in on the



Growing up: what was it that pushed the nascent Universe away from comfortable equilibrium after the Big Bang?

vexed issue of time's elusive arrow, he covers some familiar territory: the theory of relativity, basic cosmology, Mach's principle, quantum non-locality and the Bell inequalities, Wheeler's delayed-choice experiment and the many-universes interpretation of quantum mechanics. Greene even tackles, but does not resolve, the vexed issue of the flow of time — that most familiar of human experiences, which is bafflingly absent from physicists' descriptions of the world.

One way to think of the cosmic arrow of time is to regard the Universe as a gigantic clock slowly running down. The problem is to explain how it was wound up in the first place. Cosmologists naturally look to the Big Bang as the start of the whole show. But this immediately presents a puzzle. We have available a snapshot of the Universe shortly after its birth, in the form of a thermal map of the sky taken from satellites such as the Wilkinson Microwave Anisotropy Probe and various ground-based instruments. The data show the fading afterglow of the Big Bang, and enable cosmologists to reconstruct the details of its primordial thermodynamic state. The result is striking: just 380,000 years after its birth, the Universe looked virtually dead already, consisting of uniform hot gas spread through space in a state extremely

close to the heat death envisaged by von Helmholtz. So, if the young Universe was close to thermodynamic equilibrium then, how has it come alive again and achieved such a low-entropy state today?

Greene correctly identifies the solution of this puzzle. The key lies with gravity. This universal force seizes any tiny irregularity in the distribution of matter and amplifies it. Over billions of years, the smooth hot gases of the early Universe have become clumpy assemblages of galaxies, stars and black holes. This trend from smooth to clumpy constitutes gravity's very own arrow of time and trumps the rise in the entropy of matter. The upshot is that a Universe that starts out smooth, with matter in thermodynamic equilibrium, can end up clumpy, with matter dragged far from equilibrium. Since the time of the Big Bang, the entropy of matter has risen, but the maximum possible entropy has risen even faster, mostly due to the expansion of the Universe. Our very existence depends on the entropy gap that this has created.

The quest for the source of time's arrow can therefore be traced back, somewhat tortuously, to the fact that the Universe began in a relatively smooth state. When I worked on this problem in the 1970s, the initial smoothness had to be simply accepted as an unex-

plained initial condition. Today, however, there is a ready explanation in the form of the 'inflationary Universe' scenario, involving a burst of frenetic expansion shortly after the cosmic origin. Any pre-existing irregularities would have been smoothed away by this abrupt and huge swelling of space.

This pleasing conclusion immediately begs the question of what preceded inflation and why it happened at all — which provides a convenient jumping-off point for Greene to embark on a discussion of the physics of unification. This is familiar territory for him, as he is first and foremost a string theorist. He treats the reader to a review of the current state of string/M theory, with digressions into a bewildering array of topics including higher dimensions, brane worlds, the holographic paradigm, teleportation, wormholes and time travel. If I have a criticism of this book, it is that it packs in so many challenging and abstract topics that it can leave the reader's mind reeling. Sensibly, Greene repeatedly cautions that most of the ideas he gallops across are extremely speculative and are unlikely to be tested experimentally any time soon.

The nature of space and time has exercised the minds of the world's greatest scientists and philosophers for centuries. Mostly they have regarded space and time as simply there — a given. Modern physics has shown that they are, in fact, things, just as particles of matter are things. The heady developments described in this book hold out the promise that we may one day explain how space and time have come to exist, and why they possess the properties that they do. ■

Paul Davies is at the Australian Centre for Astrobiology, Macquarie University, Sydney, New South Wales 2109, Australia. His latest book is How to Build a Time Machine.

Touring artificial minds

Alan Turing: Life and Legacy of a Great Thinker

edited by Christof Teuscher

Springer: 2004. 542pp. £46, \$69.95, €59.95

John L. Casti

In 1999, *Time* magazine made Albert Einstein its 'man of the century' for the work that changed our view of time and space. It's difficult to argue too strenuously with this choice. But when it comes to scientists who have affected daily life, a far better choice would be Alan Turing, the British mathematician turned computer scientist whose invention in 1936 of what is now called the Turing machine was the theoretical backbone for every one of the zillions of computers in use today. Not only did Tur-

ing provide the key theoretical element for computing machines, he helped to build one of the first electronic computers, in Manchester, UK, shortly after the Second World War.

This book is the outgrowth of a workshop held in Lausanne, Switzerland, in June 2002 to honour the ninetieth anniversary of Turing's birth on 23 June 1912. Turing's work was so broad and deep that another gathering this year, to mark the fiftieth anniversary of his death, would not be out of place.

It is difficult to find the superlatives to describe the wonderful job the contributors to this book have done. Every chapter is written in an expository fashion, demanding very little in the way of background knowledge from any scientifically minded reader. The range of topics is also impressive, with sections on Turing's life and thoughts, the theory of computation and the Turing machine, artificial intelligence and the Turing test, the wartime Enigma code-breaking work and, finally, forgotten ideas.

Each section contains between two and seven chapters that explore themes ranging from what Turing might have thought about today's work in 'hypercomputation' — a field that explores information processing beyond the abilities of Turing machines — to his ideas on thinking machines and robots. The book's contributors are as sterling a collection of computer scientists, philosophers, engineers and historians as one could ever wish for, including logician Martin Davis, philosophers Daniel Dennett and Jack Copeland, technologist Ray Kurzweil and historian Andrew Hodges.

In her extremely entertaining chapter 'Alan's Apple: Hacking the Turing Test', the Italian writer and theatre director Valeria Patera creates a theatrical setting in which eminent figures in artificial intelligence meet in a virtual plane to consider Turing's ideas on thinking machines. A staging of this might be more interesting, intellectually at least, than the rather dull play *Breaking the Code* that ran so successfully in London and New York some years back.

Two of the more provocative contributions come from Davis, who argues against the ideas put forth by a number of researchers for transcending the Turing barrier in computation, and from Kurzweil, who explains in detail his well-known arguments



Machine architect: Alan Turing.

for why technological progress will occur at such a pace that machine intelligence will surpass the human variety within a few decades.

On the principle that no book is perfect, I have to admit to one small quibble. Given Turing's great interest in biological processes, especially near the end of his life, and his pioneering work on what we now call mathematical biology, I was disappointed to see only one of the 20 chapters devoted to that aspect of his work. Of course, no book can do everything, and this short-changing of biology in favour of com-

puting is more of an opportunity than a problem. Nevertheless, a couple more chapters on morphogenesis, artificial life and so on would have really made this book the definitive volume on Turing's work and its implications.

I unreservedly recommend this book to anyone even slightly interested in the continuing role of Turing's work in the development of computer science in particular, and ideas in general. Conference proceedings rarely make for good reading and are generally strange beasts to review. This volume is the exception that proves that rule. ■

John L. Casti is at Complexica, Santa Fe, New Mexico 87505, USA, and the Institute for Monetary Economics, Vienna, Austria.

In Newton's long shadow

From Newton to Hawking: A History of Cambridge University's Lucasian Professors of Mathematics

edited by Kevin Knox & Richard Noakes
Cambridge University Press: 2003. 512 pp.
£27.50, \$45

Lewis Pyenson

For more than 300 years, Cambridge University's Lucasian professors have promoted the mathematical elaboration of nature's laws. The founder of the chair, Henry Lucas, who was awarded an honorary MA by the university in 1636 and who represented Cambridge in both the Short and the Long Parliaments, filled a gap in his university by endowing a professorship in mathematics,

Science in culture

A weighty issue

Eduardo Chillida's sculptures are a form of 'rebellion' against Newton.

Stefano Grillo

Spanish artist Eduardo Chillida (1924–2002) used the name *Gravitation* for a major series of works that occupied him over a span of nearly 15 years—a clear suggestion that he shared some of the concerns of physical scientists. Indeed, Chillida himself once declared that he used “weight in his sculpture in order to rebel against Newton”. Was this simply a fashionably extravagant statement?

No one would claim that Chillida's sculptures and reliefs supersede Isaac Newton's theory of gravitation in the sense that Einstein's general theory of relativity does. Chillida's aim was never to create a new quantitative model to compete with that of Newton. But his statement acquires significance when seen in terms of his work's visual approach to the experience and understanding of nature. This approach consists, in the words of the Nobel-prizewinning poet Octavio Paz, of a “qualitative physics”, springing from a “direct, dynamic and non-quantitative vision of reality”.

Chillida's works allow one to visually grasp phenomena such as weight, and even its opposite, weightlessness or levitation. He achieves this chiefly through his sense of form and of space. His rhythmically twisting shapes, often supported in ways that defy our visual expectations, seem to possess a weight that is modulated by their form and is quite independent of the measurable properties of the piece. Thus Chillida's great steel and concrete structures often appear to be floating in the surrounding space, whereas the paper reliefs of his *Gravitation* series communicate a remarkable sense of visual consistency.

In the steel sculpture shown here, *De Música III*—one of 45 being exhibited at the Yorkshire Sculpture Park in Wakefield, UK, until 4 May (www.ysp.co.uk)—the horizontal structure hovers above the ground in an apparent state of levitation. Our immediate reaction is surprise that there are no additional legs to provide convincing support. The realization of the structure was a remarkable creative and technical achievement. Chillida spent



Up in the air: the floating form of *De Música III* challenges our visual expectations.

several months welding together the unsupported horizontal sections to ensure that the joints were stable enough.

The importance of this visual and non-quantitative way of experiencing reality has been stressed in recent years by the mathematician René Thom, a Fields medallist and admirer of Chillida's work. Thom noted that there are contexts in which nuclear physicist Ernest Rutherford's claim that “qualitative is nothing but poor quantitative” does not hold true. Thom's own field of topology, for example, essentially relies on qualitative distinctions.

It is also significant that Chillida provided the lithocollages to illustrate the book *Die Kunst und*

der Raum by the philosopher Martin Heidegger, who was concerned throughout his life with the foundations of modern science and with the question of whether the mathematical approach is the only valid way of understanding physical reality.

So despite Chillida's declared rebellion against newtonian physics, his work has been concerned with grasping the same phenomena in the natural world that have always interested scientists. And the result of his approach is both strikingly spectacular and thought-provoking.

Stefano Emilio Grillo is a physicist at the University of Perpignan and IMP-CNRS, Rambla de la Thermodynamique, Tecnosud, 66100 Perpignan, France.

explicitly beyond the reach of Cambridge's influential colleges.

The Lucasian professor was to lecture for one hour each week, deposit a transcript in the library, and hold office for a further two hours. It may have been at the request of the second incumbent, Isaac Newton, that King Charles II amended the statutes to allow a professor to hold a concurrent college fellowship and to require “all undergraduates past the second year and all Bachelors of Arts up to the third year” to attend the chair's lectures. The first Lucasian professor, Isaac Barrow, vacated the chair in favour of his pupil Newton, who has cast a long shadow over his 15 successors.

The sociologist Max Weber observed that, at least for the papal succession, the second-best candidate generally wins elected office. For much of the history of the Lucasian chair, even second best would be a stretch. Talent circulated freely in eighteenth-century Europe, but the Hanoverian regents hardly gave so much as a thought to looking for a Lucasian professor at Georg-August University in Göttingen or, closer to hand, in the great intellectual reservoir of Scotland, home to subtle minds from Colin Maclaurin to the historian and mathematician Thomas Carlyle.

Foreign musicians George Frederick Handel and Joseph Haydn created extraordi-

nary scores in England. A. W. Hofmann brought organic chemistry to London from Germany. Any one of the Bernoullis, Leonhard Euler or Carl Friedrich Gauss would have dramatically changed the course of history had the Lucasian electors been passionate about promoting mathematical talent. Not at Cambridge. The Lucasian professors have all been English or Irish Protestant—even the one foreign national, Paul Dirac, a naturalized Briton when he received the chair in 1932, was born in Bristol.

Carlyle described the eleventh Lucasian professor, Charles Babbage, as “a mixture of craven terror and venomous-looking vehemence”. Indeed, there is a quirky quality to

many of the incumbents, whether they espoused apostate creeds (Arianism or spiritism) or voiced unappealing prejudices (denigration of women, contempt for labouring Britons, denial of Irish political aspirations, or flirtation with Stalinism). Certainly until the nineteenth century, Cambridge mathematicians, like much else at the university, were sexually deviant; among the early Lucasian professors, only the blind John Saunderson seems to have married. There are tantalizing suggestions about sexual orientation: Newton's avoidance of the company of women, Saunderson's enjoyment of them, John Colson's engagement of his sister as a housekeeper, and Joseph Larmor's pleasure in the college baths. Perhaps mathematical equations came to the professors as compensation for repressed libido.

The Lucasian chair is a gauge of intellectual life at Cambridge. In the century or so after Newton's tenure, the Lucasian professors were minor mathematicians and inconstant discoverers of natural law, mirroring an intellectual trough in science in England during the eighteenth century. The collective obscurity of their origins matches the extent to which Cambridge was open to all intellectually promising young men. In this time, religious scripture and orthodox dogma were a burden of the chair. Going to the university meant travelling the high road to an Establishment sinecure — a church living.

The first Industrial Revolution had an impact on the Lucasian professors, leading to Babbage's tracts on political economy and his phantom calculating machine, handsomely funded from the public purse but never constructed in his time. At the height of the second Industrial Revolution, Larmor pinned down electromagnetism. By this time, the Lucasian professors, along with an increasing proportion of undergraduates, came from relatively privileged social strata.

In their solid and engaging introduction to *From Newton to Hawking*, editors Kevin Knox and Richard Noakes contend that from around 1830 "the Lucasian professors played important roles in making Britain the preeminent scientific state". But in the area where the editors' contention for eminence seems most convincing, natural history, the longest-serving chair, George Gabriel Stokes, held the great innovator Charles Darwin to be both wrong and dangerous.

Only in the later part of the twentieth century was the Lucasian chair cut free from reli-



Added value: both Isaac Newton (above) and Stephen Hawking have sought to harmonize nature and mathematics as Lucasian professors.

gious orthodoxy and social convention. With the most recent professors, Dirac, M. James Lighthill and Stephen Hawking, England once again radiates enlightenment about the mathematical harmonies of the natural world.

From Newton to Hawking also speaks to enlightened scholarship in England. It contains almost no trace of postmodernist whimsy. Each of the chapters focuses on one

professor or on a short sequence of them. By this organizational device, the book signals a return to conventional biography, the elegant prose making relatively little appeal to social statistics and psychology. But I found myself wanting to know more about the professors' income from the chair and other sources. Historical taste aside, the volume is striking for both its narrative and its original research. The Lucasian chair confirms the inertia of privileged institutions: no amount of swinishness, no succession of bad choices, no penury of emoluments, no long-running infelicities in pedagogical norms, no inattention to training acolytes, and no egregious prejudices against men and women of talent can permanently injure an institution that continues to offer positions of power and prestige to its graduates. It is a story to comfort any number of vice-chancellors, provosts, deans and department heads. ■

Lewis Pyenson is at the Center for Louisiana Studies, University of Louisiana at Lafayette, Louisiana 70504-0831, USA.

From 'not wrong' to (maybe) right

How ignoring glaring problems can sometimes lead to fruitful theories.

Frank Wilczek

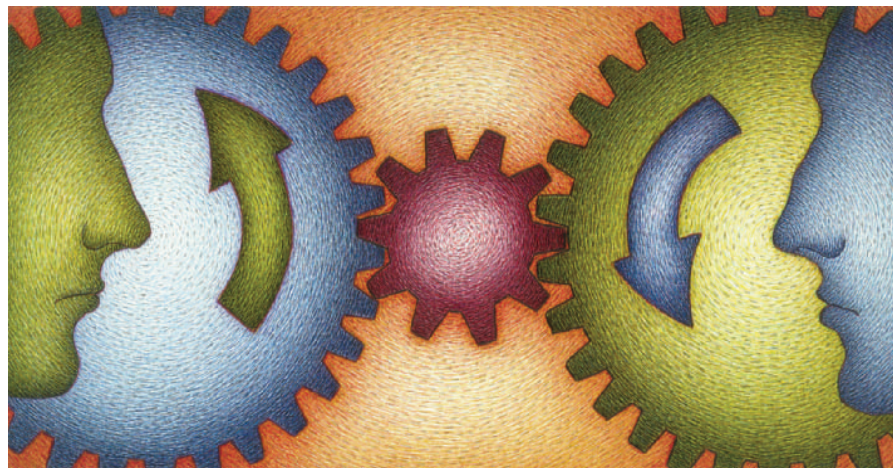
Savas Dimopoulos is always enthusiastic about something, and in the spring of 1981 it was supersymmetry. He was visiting the new Institute for Theoretical Physics in Santa Barbara, which I had recently joined. We hit it off immediately — he was bursting with wild ideas, and I liked to stretch my mind by trying to take them seriously.

Supersymmetry was (and is) a beautiful mathematical idea. The problem with applying supersymmetry is that it is too good for this world. We simply do not find particles of the sort it predicts. We do not, for example, see particles with the same charge and mass as electrons, but a different amount of spin.

However, symmetry principles that might help to unify fundamental physics are hard to come by, so theoretical physicists will not give up on them easily. Based on previous experience with other forms of symmetry, we have developed a fallback strategy, called spontaneous symmetry breaking. In this approach, we postulate that the fundamental equations of physics have the symmetry, but the stable solutions of these equations do not. The classic example of this phenomenon occurs in an ordinary magnet. In the basic equations that describe the physics of a lump of iron, any direction is equivalent to any other, but the lump becomes a magnet with some definite north-seeking pole.

Understanding the possibilities for spontaneously broken supersymmetry requires model building — the creative activity of proposing candidate equations and analysing their consequences. Building models with spontaneously broken supersymmetry that are consistent with everything else we know about physics is a difficult business. Even if you manage to get the symmetry to break, the extra particles are still there (just heavier) and cause various mischief. I briefly tried my hand at model building when supersymmetry was first developed in the mid-1970s, but after some simple attempts failed miserably, I gave up.

Savas was a much more naturally gifted model-builder, in two crucial respects: he did not insist on simplicity, and he did not give up. When I identified a particular difficulty (let us call it A) that was not addressed in his model, he would say: "It's not a real problem, I'm sure I can solve it," and the next afternoon he would come in with a more elaborate model that solved difficulty A. But then we would discuss difficulty B, and he would solve that one with a completely



different complicated model. To solve both A and B, you had to join the two models, and so on to difficulty C, and soon things got incredibly complicated. Working through the details, we would find some flaw. Then the next day Savas would come in, very excited and happy, with an even more complicated model that fixed yesterday's flaw. Eventually we eliminated all flaws, using the method of proof by exhaustion — anyone, including us, who tried to analyse the model would get exhausted before they understood it well enough to find the flaws.

When I tried to write up our work for publication, there was a certain feeling of unreality and embarrassment about the complexity and arbitrariness of what we had come up with. Savas was undaunted. He even maintained that some existing ideas about unification using gauge symmetry, which to me seemed genuinely fruitful, were not really so elegant if you tried to be completely realistic and work them out in detail. In fact, he had been talking to another colleague, Stuart Raby, about trying to improve those models by adding supersymmetry! I was extremely sceptical about this 'improvement', because I was certain that the added complexity of supersymmetry would spoil the existing success of gauge symmetry in explaining the relative values of the strong, electromagnetic and weak coupling constants. The three of us decided to do the calculation, to see how bad the situation was. To get oriented and make a definite calculation, we started by doing the crudest thing, which was to ignore the whole problem of breaking supersymmetry. This allowed us to use very simple (but manifestly unrealistic) models.

The result was amazing, at least to me. The supersymmetric versions of the gauge symmetry models, although they were vastly

different from the originals, gave very nearly the same answer for the couplings.

That was the turning point. We put aside the 'not wrong' complicated models with spontaneous supersymmetry breaking, and wrote a short paper that, taken literally (with unbroken supersymmetry), was wrong. But it presented a result that was so straightforward and successful that it made the idea of putting gauge symmetry and supersymmetry unification together seem (maybe) right. We put off the problem of how supersymmetry gets broken. And today, although there are some good ideas about it, there is still no generally accepted solution.

After our initial work, more precise measurements of the couplings made it possible to distinguish between the predictions of models with and without supersymmetry. The models with supersymmetry work much better. We all eagerly await operation of the Large Hadron Collider at CERN, the European particle physics laboratory, where, if these ideas are correct, the new particles of supersymmetry — or, you might say, the new dimensions of superspace — must make their appearance.

This little episode, it seems to me, is 179 degrees or so out of phase from Karl Popper's idea that we make progress by falsifying theories. Rather, in many cases, including some of the most important, we suddenly decide our theories might be true, by realizing that we should strategically ignore glaring problems. It was a similar turning point when David Gross and I decided to propose quantum chromodynamics (QCD) based on asymptotic freedom, putting off the problem of quark confinement. But that is another story... ■

Frank Wilczek is in the Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307, USA.

RNA finds a simpler way

Thomas R. Cech

Is there no end to the versatility of RNA? The latest feat to be revealed is RNA's ability to switch off genes through a neatly straightforward mechanism. So it isn't only proteins that can repress gene activity.

Organisms profit from synthesizing only those enzymes whose products are in demand. After all, if there's plenty of an amino acid around, why should a cell waste energy making the enzymes that are needed to generate it from precursors? A standard textbook mechanism by which this is achieved is repression. Repressor proteins sense when there is a build-up of a certain product, bind to the gene that encodes the enzyme that generates this product, and thereby inhibit the transcription of more messenger RNA (mRNA; Fig. 1a). On page 281 of this issue, Winkler *et al.*¹ describe a bacterial gene-regulation system with a twist. The repressor is not a protein, but instead is a switchable (on–off) self-cleavage element within the mRNA itself.

The newly discovered molecular switch involves an RNA molecule with enzymatic activity. Self-cleavage by this ribozyme is accelerated 1,000-fold in the presence of glucosamine-6-phosphate (GlcN6P), a small sugar. GlcN6P is generated by the GlmS enzyme, which is encoded by a portion of the *glmS* mRNA downstream from the ribozyme sequence. So it is easy to envisage a gene-regulatory circuit in which the *glmS* mRNA is translated into GlmS protein until the GlcN6P product accumulates; at that point, GlcN6P binds to the special catalytic element in the mRNA, causing it to self-destruct (Fig. 1b). Although the cleavage event itself leaves the coding region of the message intact, the RNA is either destabilized and subject to degradation, or its ability to be translated into functional protein is compromised.

The minimal region of the RNA that can confer this regulatory activity is roughly 75 nucleotides long, the size of a transfer RNA. The chemical mechanism by which it is cleaved during repression has previously been seen in other ribozymes, such as the hammerhead, hairpin and hepatitis delta virus ribozymes², but the structure of its catalytic fold appears to be distinct from these. When placed upstream of an unrelated 'reporter' gene, the *glmS* ribozyme element also repressed its expression, and this required the same sequences that are required for *glmS* self-cleavage *in vitro*¹. Thus, the active RNA element is modular and transplantable.

Although switching gene expression in this manner is new, there are precedents for the individual components of this regulatory

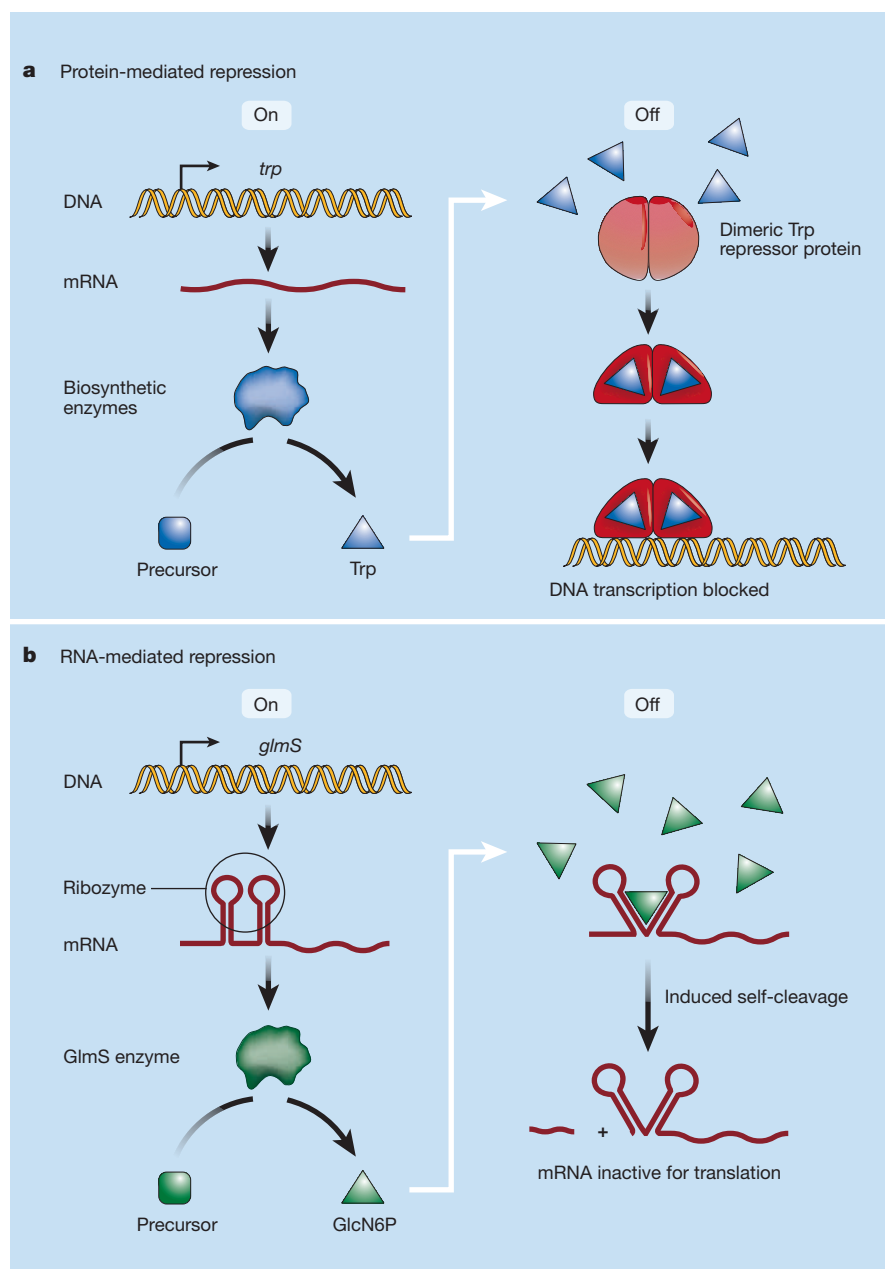


Figure 1 Repression systems. **a**, Classic repression of the genetic unit involved in production of the amino acid tryptophan. Transcription of the DNA encoding the biosynthetic enzymes for tryptophan produces messenger RNA, which is translated to make the enzymes themselves. When enough tryptophan (blue triangles) has been produced, the tryptophan binds to the dimeric Trp repressor protein, inducing a conformational change that enables it to bind the DNA, thereby blocking transcription. **b**, Repression of a gene at the RNA level by metabolite-activated ribozyme cleavage, as described by Winkler *et al.*¹. The GlmS enzyme, which is involved in the synthesis of GlcN6P (green triangle), is encoded by the *glmS* mRNA. Also in this RNA is a ribozyme sequence. In the presence of GlcN6P, the self-cleavage activity of the ribozyme is markedly enhanced; ribozyme self-destruction renders the *glmS* mRNA non-functional, thereby inhibiting further production of GlcN6P.

circuit. First, consider GlcN6P binding. Even though it once seemed unlikely that RNA, with its limited diversity of chemical groups and its high negative charge, could specifically bind small molecules, we now know that some naturally occurring RNAs do just that. For example, a particular group of ribozymes forms a pocket that binds guanosine monophosphate, one of the four monomer building-blocks of RNA³. And a specific region of RNA from the human immunodeficiency virus binds a derivative of the amino acid arginine⁴. More recently, short (<100 nucleotides) RNA 'aptamers' have been identified that specifically bind everything from hydrophobic amino acids to small organic molecules to metal ions^{5,6}. In terms of specificity, an RNA aptamer can even distinguish the plant alkaloid theophylline from the closely related molecule caffeine⁷.

As a second precedent, aptamers found within some natural mRNAs bind small molecules as part of gene-regulatory feedback circuits. In the bacterium *Escherichia coli*, coenzyme B12 binds directly to, and thereby represses translation of, the mRNA coding for the protein that transports its precursor, cobalamin⁸. In *Bacillus* species, the synthesis of thiamin and riboflavin involves discrete genetic units. These operons are controlled by direct binding of thiamin pyrophosphate and flavin mononucleotide to leader sequences of the corresponding mRNAs, resulting in transcription coming to an early finish⁹. Although structured RNAs had previously been found to regulate gene expression at multiple levels — during gene transcription, splicing of the RNA and the translation of mRNA into protein — these new systems are different in that they consist entirely of RNA. No protein seems to be required for either sensing the concentration of the metabolic end-product or switching off gene expression.

Finally, several groups had previously engineered artificial riboswitches — RNA aptamers containing ribozyme sequences that, on binding small molecules, induce ribozyme-mediated cleavage of the RNA¹⁰. These findings might have prompted some to wonder why nature failed to use, or perhaps to retain, such an elegant mechanism. But it's clear now that this perhaps-ancient talent of RNA is alive and well, and is currently used by at least some bacteria to control their *glmS* genes. An exciting question is how widespread these control elements are in biology.

In their eloquent and prescient article on genetic regulatory mechanisms, François Jacob and Jacques Monod¹¹ thought that genetic repressors might be RNA. They nonetheless much preferred the idea of a protein repressor, as, in their own words "the capacity to form stereospecific complexes with small molecules appears to be a privilege of proteins". When many repressors were subsequently identified as proteins, the idea

that RNA repressors existed faded away. But more recently, the concept that RNA can in fact bind small molecules with high specificity and affinity has been established. In a sense, the recent discoveries of naturally occurring 'RNA biosensors' have gone full-circle back to Jacob and Monod — RNA can be the active element that switches off repressible genes in response to the concentration of a cellular metabolite. ■

Thomas R. Cech is at the Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, Maryland 20815, and in the Department of Chemistry and Biochemistry, University of

Colorado, Boulder, Colorado 80309, USA.

e-mail: thomas.cech@colorado.edu

1. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. *Nature* **428**, 281–286 (2004).
2. Doudna, J. A. & Cech, T. R. *Nature* **418**, 222–228 (2002).
3. Michel, F., Hanna, M., Green, R., Bartel, D. P. & Szostak, J. W. *Nature* **347**, 578–580 (1989).
4. Puglisi, J. D. *et al.* *Science* **257**, 76–80 (1992).
5. Ellington, A. D. & Szostak, J. W. *Nature* **346**, 818–822 (1990).
6. Gold, L., Polisky, B., Uhlenbeck, O. & Yarus, M. *Annu. Rev. Biochem.* **64**, 763–797 (1995).
7. Zimmermann, G. R., Jenison, R. D., Wick, C. L., Simorre, J. P. & Pardi, A. *Nature Struct. Biol.* **4**, 644–649 (1997).
8. Nahvi, A. *et al.* *Chem. Biol.* **9**, 1043–1049 (2002).
9. Mironov, A. S. *et al.* *Cell* **111**, 747–756 (2002).
10. Silverman, S. K. *RNA* **9**, 377–383 (2003).
11. Jacob, F. & Monod, J. *J. Mol. Biol.* **3**, 318–356 (1961).

Astronomy

We can see clearly now...

Nicholas White

The mystery of the diffuse γ -ray glow that pervades the Milky Way has been solved, thanks to a space telescope with the power to resolve compact γ -ray sources in the Galaxy.

The Milky Way glows brightly at γ -ray wavelengths. Although this glow was discovered more than three decades ago, its origin has been a mystery. Now the Milky Way has been observed at these wavelengths with new clarity, by astronomers using the European Space Agency (ESA) INTEGRAL observatory. In this issue (*Nature* **428**, 293–296; 2004), F. Lebrun *et al.* report that the soft γ -ray glow is the sum of radiation from previously unresolved point sources, many of them black holes and neutron stars, buried in gas and dust. This discovery points to a future in which soft γ -rays will become a powerful tool for finding black holes in otherwise obscured regions.

Radiation such as γ -rays and their less energetic cousins, X-rays, is used in everyday applications to see inside objects — for example, in medical scanners and for security screening at airports. This same penetrating power makes them a useful astronomical probe for finding and studying objects that may be buried deep in obscuring clouds of dust and gas. By the same token, the capability of γ -rays to penetrate rather than be reflected makes it challenging to design γ -ray telescopes — until recently these telescopes have been little more than crude light buckets, with poor angular resolution (a few degrees, equivalent to many times the diameter of the Moon).

Despite their penetrating power, γ -rays are eventually stopped and absorbed in the extreme depth of Earth's atmosphere — which is fortunate for life on Earth! Consequently, the γ -ray window is closed to Earth-bound telescopes; only with the start of the space age was it opened up for astronomical exploration. The first γ -ray telescopes, launched in the 1970s, detected many bright cosmic sources,

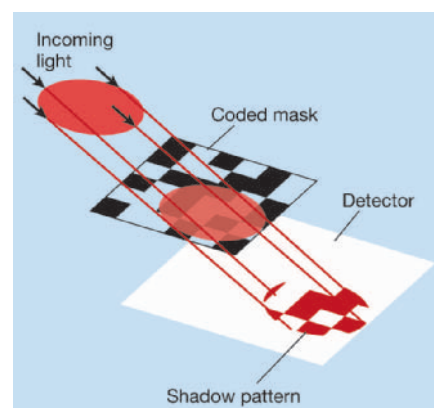


Figure 1 The coded-mask technique. In optical telescopes, mirrors or lenses are used to focus light, but these same lenses would absorb γ -rays. The coded-mask technique is similar to the principle of a pinhole camera. The mask partly covers the opening of the telescope and casts a shadow on the detector below. Because the position of the shadow depends on the orientation of the mask with respect to the source, it is possible to determine the position and intensity of the source (or sources) and thus produce an image of the γ -ray sky.

including the γ -ray glow of the Milky Way.

The origin of these soft γ -rays was puzzling. If it is truly a diffuse glow, then the energy being released is huge and would have ionized the molecules in the interstellar medium. The most promising explanation was that this emission originates from an unknown population of point sources, which is spread out on the sky but separated by less than the angular resolution of the γ -ray telescopes available at the time. Confirming this theory required a new large telescope with much sharper imaging.

To achieve the necessary sharper imaging, astronomers have resorted to a 'shadow-gram' technique, using a metal mask encoded with a pattern of holes to cast a shadow on a large camera (Fig. 1). By correlating the observed shadow with the known pattern in the coded mask, the position of the γ -ray source can be pinpointed. The viability of the technique was demonstrated by a pathfinder Russian–French telescope that flew on the Russian GRANAT mission in 1989.

Coded-mask telescopes need large radiation-collecting areas to detect the faint sources, and this in turn requires a big, heavy space observatory. For more than two decades, astronomers could only dream of the potential of such a γ -ray telescope for solving the mystery of our Galaxy's γ -ray glow. ESA took up the challenge in the mid-1990s and led the development of the INTEGRAL observatory. In October 2002, it was launched on a Russian Proton rocket.

The INTEGRAL observations reported by Lebrun *et al.* reveal many dozens of sources, most of them quite faint. The sum of their γ -radiation matches the γ -ray flux previously attributed to the diffuse glow. Many of the newly discovered sources seem to be either black holes or neutron stars.

They show absorption spectra — corresponding to their radiation being absorbed by some intervening material between these sources and INTEGRAL, meaning that they stand out only at γ -ray wavelengths. The absorbing material may be either the accretion flow from a companion star, or a molecular cloud in which the source is embedded. Now we know where these objects are, follow-up observations using other telescopes, such as ESA's XMM–Newton Observatory, will uncover their nature more fully.

Over the coming years, INTEGRAL will make a complete census of the buried populations of black holes and neutron stars in the Milky Way. Exploiting its capabilities to the full, the observatory will be able to search nearby galaxies for supermassive black holes buried in their centres, pick up γ -ray bursts and detect the radioactive decay of newborn elements in the fires of exploding supernovae. The Universe will appear to be a very different place when viewed clearly with γ -ray eyes. ■

Nicholas White is at the Laboratory for High Energy Astrophysics, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.
e-mail: nicholas.e.white@nasa.gov



100 YEARS AGO

I happened to possess a small sample of a red deposit, coloured by iron, which is left by the water of the King's Spring, at Bath. It occurred to me that it might be worth while to test this for radio-activity. The result was to show that the deposit was markedly active. On leaving it in the testing vessel (which was closed airtight) for a few days, the activity was found to increase to several times its initial value. This shows that the deposit gives off an emanation freely, even without heat. Experiments were then made to test the rate of decay of this emanation. It proved to be identical with the rate of decay of the emanation of radium... The presence of radium in the Bath water and deposits is of special interest because of the occurrence of helium in the gas which arises from the spring... There can be little doubt that the helium owes its origin to the same store of radium that supplies the water.
From *Nature* 17 March 1904.

50 YEARS AGO

At a meeting of the Geological Society held at Burlington house on February 24, Mr. A. T. Marston exhibited a number of flints which he had attempted to stain with chromic acid or potassium dichromate in order to assist appreciation of the condition of the Piltdown flint recently described by Drs. Oakley and Weiner (*Nature*, December 12, 1953). Mr. Marston pointed out that though some types of these flints would not stain at all, others became deeply stained. In the latter case, however, the stain was very difficult to remove. As the stain on the Piltdown flint had apparently been easily removed by acid, Mr. Marston queried whether it had indeed been artificially produced, and also inquired whether there was any possibility that traces of chromium might exist naturally in the ferruginous Piltdown gravels. Dr. Oakley then exhibited specimens of flint from Piltdown, and described tests that had been made on all the flints illustrated in the original paper by C. Dawson and A. Smith Woodward. All these flints had a ferruginous stain which was easily removed by acid, in contrast to the other 'natural' flints from the Piltdown area collected at the same time... Dr. Oakley believed that the illustrated flints had been stained artificially with a ferruginous solution, and in one case a chromic solution had also been used in order to make the flint less red, and so resemble more closely the 'natural' flints.

From *Nature* 20 March 1954.

Cell biology

The strain of being a prion

Mick F. Tuite

Prions are remarkable infectious agents associated with certain brain diseases. But they also occur in fungi, experiments with which now provide plausible answers to some critical questions about prion biology.

A widely (but not universally) accepted dogma about the agents known as prions is that they are protein-based entities that are self-perpetuating, 'infectious' and devoid of any transmissible nucleic acids. Yet despite intensive research into this 'protein only' hypothesis¹, two crucial challenges have remained unanswered. Can infectivity with purified prion protein be demonstrated? And how can different prion 'strains' be generated without any underlying change in the amino-acid sequence of the prion protein or in the genetic make-up of the host? Papers in this issue by King and Diaz-Avalos² and Tanaka *et al.*³ (pages 319 and 323) address these challenges. They provide the most dramatic demonstration to date of the validity of the protein-only hypothesis.

Infection of the host by a prion can, over time, lead to the replication and subsequent transmission of an aggregated fibrous form of the infectious protein known as an amyloid. In mammals, the only known prion protein (PrP) is associated with a group of neurodegenerative diseases that include Creutzfeldt–Jakob disease in humans and

bovine spongiform encephalopathy in cows⁴. Prion proteins have also been linked with stable, heritable traits in two fungi (*Saccharomyces cerevisiae* and *Podospora anserina*), although none of the four known fungal prions can be considered 'disease-causing' — some may in fact be beneficial to the host⁵.

King and Diaz-Avalos², and Tanaka *et al.*³, have exploited the prion properties of the Sup35p protein of *S. cerevisiae*. Sup35p is an essential factor required by the protein-synthesizing organelles — the ribosomes — to terminate synthesis of a polypeptide chain. Cells in which most of the Sup35p protein molecules have been converted to their prion form ($[PSI^+]$ cells) are defective in this crucial termination phase.

$[PSI^+]$ cells can be readily detected in a population of 'normal' genetically marked $[psi^-]$ cells by a simple screen involving cell colour and protein analysis (Fig. 1, overleaf). Using this screen, King and Diaz-Avalos, and Tanaka *et al.*, independently developed what are essentially protein-only transformation systems and used them to demonstrate that

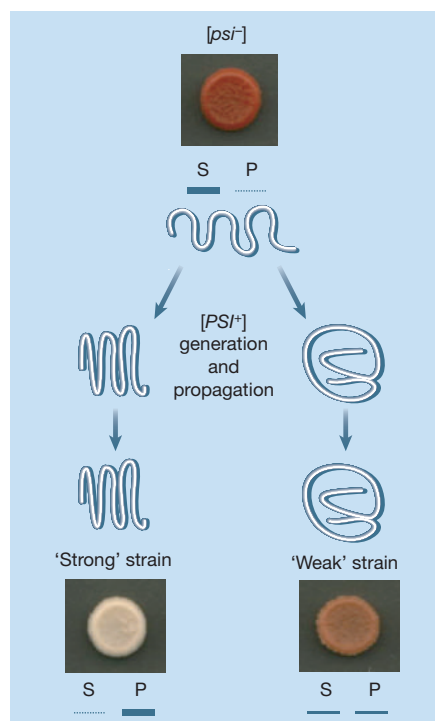


Figure 1 Prion behaviour in the yeast *S. cerevisiae*. Tell-tale colony colours are shown, along with an indication of the amount of Sup35p protein in soluble (S lane) or aggregated (P lane) form. $[psi^-]$ yeast cells contain a soluble form of the Sup35p protein. Such cells fail to suppress the *ade1-14* mutation, because of effective translation termination, leading to red colonies being formed on agar plates. Two different $[PSI^+]$ strains can be generated *de novo*; they differ not only in the conformation of Sup35p, but also in the extent to which the Sup35p protein is aggregated. In the 'strong' $[PSI^+]$ strain, more than 95% of the Sup35p protein is in the non-functional, aggregated fraction (lane P). These cells produce white colonies because translation termination is impaired and the *ade1-14* mutation is suppressed. In the 'weak' $[PSI^+]$ strain, a significant level of Sup35p is also seen in the soluble fraction (lane S). The 'weak' strain therefore shows a partial restoration of translation termination and concomitantly less efficient suppression of the *ade1-14* mutation giving pink colonies. Once a particular strain has been generated, it is stably replicated in that conformation.

$[psi^-]$ cells can be 'infected' — that is, turned into $[PSI^+]$ cells — with an amyloid form of part of the Sup35p protein. The part they use is one end of the protein, the N region, which contains all the structural information needed for maintaining the $[PSI^+]$ state *in vivo* and can readily assume an amyloid form *in vitro*⁵. By generating amyloid forms of the protein that are free from any other yeast protein, and by showing that the resulting infectivity is not affected by treatment with agents that destroy nucleic acids, the authors' findings leave little doubt that a single protein species can act as an infectious agent based on its conformation. Moreover, previous work with Sup35p in *S. cerevisiae*⁶ and with a different fungal prion protein (the HET-s protein of *P. anserina*)⁷, did indeed provide some evidence in favour of the view that it is the amyloid form of the protein that transmits infectivity⁸. That evidence fell short of being compelling, but the gap is now filled by the new results^{2,3}.

The existence of different prion 'strains' has always cast a shadow on the protein-only hypothesis. In mice, for example, at least 20 prion strains have been described which produce different disease characteristics but are not the result of a change in host genetic make-up⁹. It has been proposed that prion strains could arise through the existence of distinct, self-propagating conformers of otherwise identical PrP polypeptide chains. But no direct experimental evidence was forthcoming⁹ — until now.

There are at least two distinct $[PSI^+]$ strains (or 'variants'), as defined by the degree of suppression of a host nonsense mutation (*ade1-14*), and biochemically by the proportion of Sup35p that remains soluble and hence free to participate in the termination

process (Fig. 1). During the *de novo* appearance of the $[PSI^+]$ prion, both types of strain can emerge within a population of otherwise genetically identical $[psi^-]$ cells. Once a prion strain has emerged, it is stably propagated and efficiently transmitted to daughter cells during cell division. Evidence already exists that Sup35p may be in a different conformational state in different strains; for example, Sup35p from different $[PSI^+]$ strains has different amyloid seeding propensities *in vitro*¹⁰.

King and Diaz-Avalos², and Tanaka *et al.*³, provide the first direct evidence that conformationally distinct, amyloid forms of Sup35p underlie the transmissible differences in the $[PSI^+]$ strains. They show that at least two distinct conformers of Sup35p can be generated *in vitro* as defined by various biophysical measurements. By demonstrating that the 'infection' of $[psi^-]$ cells with the

different conformers each generates different $[PSI^+]$ variants, both groups firmly establish the link. Furthermore, these conformational differences are propagated *in vivo*³.

At least for yeast prions, attention can now turn to the baffling matter of how the amyloid form is propagated, thereby ensuring stable transmission to daughter cells. Two basic mechanisms, known as template-directed refolding and seeded nucleation, have been widely touted¹¹. But what is the molecular nature of the propagating unit (or seed)? Is it a conformationally altered monomeric form of the protein, or the highly aggregated amyloid form? These questions can now be tackled using the new prion transformation assays, and there are already candidates. For example, high-molecular-weight Sup35p aggregates in $[PSI^+]$ cells can be broken down into smaller, detergent-insoluble Sup35p-based polymers, and the size of the polymer depends on the $[PSI^+]$ strain¹². Could these detergent-insoluble polymers be the active seeds? We now have a powerful way of approaching the remaining mysteries of prion propagation in yeast, and one that could equally be applied to the cells of higher organisms.

Mick F. Tuite is in the Department of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK. e-mail: M.F.Tuite@kent.ac.uk

1. Prusiner, S. B. *Science* **216**, 136–144 (1982).
2. King, C.-Y. & Diaz-Avalos, R. *Nature* **428**, 319–323 (2004).
3. Tanaka, M., Chien, P., Naber, N., Cooke, R. & Weissman, J. S. *Nature* **428**, 323–328 (2004).
4. Collinge, J. *Annu. Rev. Neurosci.* **24**, 519–550 (2001).
5. Uptain, S. M. & Lindquist, S. L. *Annu. Rev. Microbiol.* **56**, 703–741 (2002).
6. Sparrer, H. E., Santos, A., Szoka, F. & Weissman, J. S. *Science* **289**, 595–599 (2000).
7. Maddelein, M. L., Reis, S. D., Duvezin-Caubet, S., Coulay-Salin, B. & Saupe, S. J. *Proc. Natl Acad. Sci. USA* **99**, 7402–7407 (2002).
8. Liebman, S. W. *Proc. Natl Acad. Sci. USA* **99**, 9098–9100 (2002).
9. Bruce, M. E. *Br. Med. Bull.* **66**, 99–108 (2003).
10. Uptain, S. M., Sawicki, G. J., Caughey, B. & Lindquist, S. *EMBO J.* **20**, 6236–6245 (2001).
11. Tuite, M. F. & Cox, B. S. *Nature Rev. Mol. Cell Biol.* **4**, 878–889 (2003).
12. Kryndushkin, D. S., Alexandrov, I. M., Ter-Avanesyan, M. D. & Kushnir, V. V. *J. Biol. Chem.* **278**, 49636–49643 (2003).

Cancer

Survival pathways meet their end

Frank McCormick

Conventional chemotherapeutic approaches to treating tumours can be hit-and-miss. One way to ensure successful treatment may be to go for the jugular of cancer-cell survival signalling as well.

Chemotherapy uses powerful drugs designed to induce cancer cells to commit suicide. So why don't all tumours succumb to these drugs? The answer is that cancer cells have an inbuilt urge to survive, so any genetic change that favours survival amidst adverse conditions will be selected for. The outcome, of course,

is that some tumours survive exposure to even the most potent therapeutic agents. But by increasing our knowledge of the components involved in the pathways that mediate survival, the hope is that targeting specific molecules will impart sensitivity to chemotherapy — a combinatorial approach. Wendel and co-workers, reporting

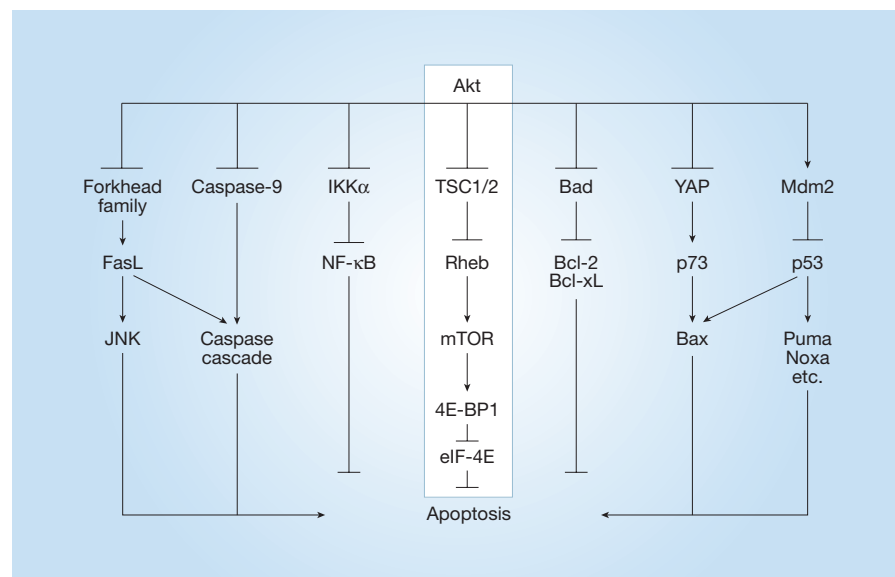


Figure 1 The complexity of survival through Akt signalling. There are at least seven signalling pathways downstream of Akt, all containing several protein components, as indicated here. Mainly by inhibiting signals that induce cell death, Akt promotes survival. An important branch of the Akt pathway for cancer-cell survival seems to be that involving the kinase mTOR. Phosphorylation by Akt of TSC proteins renders them unstable, thereby relieving, through Rheb, the inhibitory constraint on mTOR. This in turn allows mTOR to target 4E-BP1, which then suppresses eIF-4E. The ultimate consequence is suppression of cell death.

on page 332 of this issue¹, make major progress towards meeting this challenge. One important survival signalling pathway is mediated by phosphoinositide 3-OH kinase (PI(3)K) and its downstream target, Akt, and Wendel *et al.* show that blocking an enzyme, mTOR, which acts further down the pathway, sensitizes tumours to killing by conventional chemotherapy agents. They have also made stalwart progress in determining the main effector of Akt in cancer-cell survival.

Efforts to pinpoint the molecular basis of Akt's survival effects have taken several directions² (Fig. 1). As a kinase, the natural tendency of Akt is to add a phosphate group to suitable substrates, and Bad and caspase-9, proteins that encourage cells to die, were among the first targets of Akt to be identified as being phosphorylated³. Then came other cell-death regulators — IKK α , the Forkhead transcription factors, Mdm2 and YAP³. So blocking Akt might be expected to induce apoptotic cell death at many levels (Fig. 1). But whether any one of these candidates can shoulder responsibility for acting downstream of Akt in cancer-cell survival *in vivo* is not clear, because each works in certain cell types or under specific conditions.

More recently, attention has shifted to a different branch of the PI(3)K–Akt pathway, in which mTOR features. mTOR was thrust into the limelight last year when a brilliant convergence of genetics and biochemistry led to the construction of this new branch⁴ (Fig. 1). mTOR is regulated by the protein Rheb. Rheb, in turn, is controlled by the tumour-suppressor TSC proteins,

which keep cell proliferation in check. The availability of mTOR inhibitors such as rapamycin meant that this arm of the pathway could be selectively blocked, an approach not yet available to other pathways downstream of Akt.

Wendel and colleagues¹ used a mouse model of B-cell lymphoma, a cancer of antibody-producing B cells, to explore the consequences of inhibiting Akt. Tumours that needed Akt to survive were refractory to conventional chemotherapy agents such as cytoxan or doxorubicin when these agents were used alone, but were markedly sensitized to these drugs when the survival effects of activated Akt were blocked with rapamycin. In contrast, tumours that depended on another anti-apoptotic protein, Bcl-2, were not sensitized to chemotherapy-induced death by rapamycin.

Rapamycin targets mTOR, but what lies downstream of mTOR? The best-known targets of mTOR are involved in making proteins from messenger RNA. Interest in one of these targets, S6 kinase, dates back to when signal transduction pioneers — Thomas, Blenis, Rosen, Maller, Avruch and Krebs — recognized that phosphorylation of the S6 protein, which is involved in protein synthesis, could be used to indicate that external growth factors were activating upstream intracellular signalling pathways. Working their way upstream of S6, the kinase responsible (S6 kinase) was identified and a pathway could, perhaps, be traced all the way back to the receptors, at the cell surface, for factors that induce growth and proliferation. In studying this pathway, it

was known that higher levels of S6 phosphorylation are a common feature of cancer cells, including those in which the Akt pathway is upregulated through loss of the negative Akt regulator PTEN⁵.

The other major target of mTOR is 4E-BP1, discovered as a negative regulator of the elongation initiation factor eIF-4E^{6,7}. The latter protein is rate limiting for the translation of many mRNAs into proteins, but this ability depends on the degree to which it is bound to 4E-BP1. So changing the degree of S6 phosphorylation or eIF-4E activity could theoretically affect the synthesis of proteins that regulate cell proliferation — in other words, the ability of rapamycin to sensitize tumours to apoptosis could be caused by its inhibiting the growth and division needed for cancerous cells to thrive.

Recent publications, however, have suggested that eIF-4E has another function — suppressing apoptosis. Apoptosis induced by the *c-myc* gene is suppressed by eIF-4E, apparently by eIF-4E inhibiting the release of cytochrome *c* from mitochondrial membranes, a point of no return in cell death⁸. Whether this is a direct effect or the result of selective translation of proteins that regulate this process remains to be seen. The notion that eIF-4E has functions that are distinct from RNA translation has also come from the discovery that it exports the mRNA for a cell-cycle regulator, cyclin D1, out of the nucleus⁹. The implications of these findings clearly remain to be revealed.

Finally, because eIF-4E lies downstream of mTOR, Wendel and co-workers¹ asked whether artificially overexpressing eIF-4E could overcome the apoptosis-sensitizing effects of rapamycin. It could, although this is hardly surprising in view of the ability of eIF-4E to mimic Akt in promoting tumour formation and suppressing apoptosis.

The clinical importance of these results is obvious. Indeed, clinical trials of rapamycin and its analogues in combination with other chemotherapeutic agents are already underway. However, it is implicit that such an approach is only likely to work in cancers that suppress apoptosis by using the mTOR–eIF-4E pathway downstream of Akt. Tumours that rely on Bcl-2, for example, for their anti-apoptotic signals, are predicted to be unresponsive. Similarly, tumours that activate eIF-4E through mTOR-independent pathways would be expected to be insensitive to rapamycin.

Wendel and colleagues' work is an excellent example of two related themes emerging for the future of cancer treatment — designer therapies and combinatorial approaches. The former takes into account the genetic make-up of the tumours and thus the molecules that are dysregulated. In the latter approach, intelligent combinations of drugs can show marked synergy.

Moving back to the lab, the next step will be to understand in molecular detail where survival pathways and chemotherapy agents intersect.

Frank McCormick is at the Comprehensive Cancer Center, University of California, San Francisco, Box 0128, 2340 Sutter Street, San Francisco, California 94143, USA.

e-mail: mccormick@cc.ucsf.edu

1. Wendel, H.-G. *et al.* *Nature* **428**, 332–337 (2004).

2. Vivanco, I. & Sawyers, C. L. *Nature Rev. Cancer* **2**, 489–501 (2002).
3. Basu, S., Totty, N. F., Irwin, M. S., Sudol, M. & Downward, J. *Mol. Cell* **11**, 11–23 (2003).
4. Manning, B. D. & Cantley, L. C. *Trends Biochem. Sci.* **28**, 573–576 (2003).
5. Neshat, M. S. *et al.* *Proc. Natl. Acad. Sci. USA* **98**, 10314–10319 (2001).
6. Lin, T. A. *et al.* *Science* **266**, 653–656 (1994).
7. Pause, A. *et al.* *Nature* **371**, 762–767 (1994).
8. Li, S. *et al.* *J. Biol. Chem.* **278**, 3015–3022 (2003).
9. Topisirovic, I. *et al.* *Mol. Cell. Biol.* **23**, 8992–9002 (2003).

Semiconductor physics

Quick-set thin films

Mercouri G. Kanatzidis

Transistors that have active components based on thin films, rather than silicon, are attractive for many applications. The latest thin-film fabrication technique has the potential for industrial-scale production.

The original working transistor, invented at Bell Labs in the 1940s, was based on semiconducting germanium and had a junction (sandwich) configuration. But by the 1960s, this design had given way to the simpler field-effect transistor — in particular, the silicon-based MOSFET (for metal-oxide-semiconductor field-effect transistor). A typical computer processor today contains around 42 million such transistors, and demand for ever-faster computers is only increasing. As a result, the market is pushing for a downsizing of transistor technology.

However, certain applications (such as flat-panel displays) require larger-area transistors than can normally be created using silicon-based devices. Thin-film semiconductors have been explored as an alternative, although with limited success. But now it seems that the large-scale, low-cost fabrication of such devices is a step closer: on page 299 of this issue, Mitzi *et al.*¹ describe a chemical-deposition method for producing uniform films of the chalcogenides SnS_2 or SnSe_2 for use in thin-film transistors (TFTs). The resulting TFTs support large current densities (more than 10^3 A cm^{-2}), and mobilities greater than $10 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ — almost ten times larger than achieved for semiconducting films formed using the spin-coating technique (in which a solution on a substrate is spun rapidly, causing the film to spread outwards).

In a TFT, the thin film (usually silicon) is the active current-carrying layer (Fig. 1). The film sits on a substrate, which is usually glass owing to its low cost, high optical transparency and compatibility with conventional semiconductor processing technology. Recently, however, plastic has emerged as a viable challenger because of its additional flexibility, although the development of TFT technology for use with plastic substrates is still in its infancy.

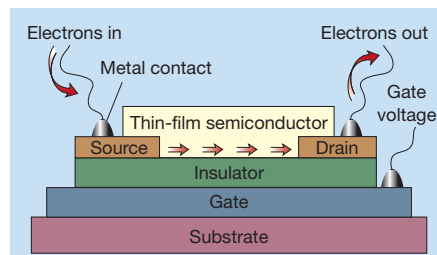


Figure 1 Cross-section of a thin-film transistor. A voltage applied at the gate controls the flow of electrons (resistance) from the source to the drain; a positive gate voltage attracts electrons to the bottom surface of the semiconductor layer and creates a conduction channel. When a voltage difference is applied between the two connector wires, electrons enter at one end (the source) and exit at the other (the drain), resulting in a current along the channel. Mitzi *et al.*¹ have now come up with a chemical-deposition method that produces uniform films of the chalcogenide SnSe_2 for use in thin-film transistors.

Transistors for high-performance display applications should have high electron mobilities, low leakage currents and low threshold voltages. But processing temperatures must also be low (below 150°C) if the transistors are to be compatible with low-cost plastic substrate materials. So the emphasis in developing large-scale TFTs has been on low-temperature deposition and the exploration of materials other than amorphous silicon. Approaches include vacuum deposition² (suitable for growing ultrathin organic films and multilayer structures), solution-deposition technologies³ (suitable for inorganic materials), and many others⁴. But these are generally not high-throughput processes. Although spin-coated semiconductor films have suffered from low mobilities^{5–7}, this technique shows much promise.

The attraction of using inorganic semiconductors lies in their stability, thermal robustness and high mobilities. The metal chalcogenides, for example, are excellent candidates for use in TFT technologies. They form a large class of compounds that are composed of one or more metals plus one of the chalcogen atoms such as sulphur, selenium or tellurium. Moreover, the energy required to delocalize a charge carrier (the energy gap)⁸ in these materials is suitable for room-temperature devices, and can be further tuned for a given application⁹.

Mitzi *et al.*¹ describe a means of creating chalcogenide active layers for TFTs through spin coating. Their continuous, uniform, ultrathin semiconducting films are only a few unit cells thick. The key to the fabrication chemistry is hydrazine (N_2H_4), which Mitzi *et al.* use as a solvent. When metal and chalcogens dissolve in hydrazine, they form chalcogenometallate solutions containing anions such as $[\text{Sn}_2\text{S}_6]^{4-}$, as well as hydrazinium cations (N_2H_5^+). These solutions can be used as precursors for spin-coating thin films of the salt $(\text{N}_2\text{H}_5)_4[\text{Sn}_2\text{S}_6]$, which then decompose to the binary metal chalcogenide at low temperature. The advantage of having hydrazinium cations, and not some other organic cations¹⁰, is that they readily and cleanly react with the counterion of $[\text{Sn}_2\text{S}_6]^{4-}$ to give continuous, crystalline semiconducting films as thin as 5 nanometres.

It is this simple chemistry that not only makes the work of Mitzi *et al.*¹ attractive, but probably technologically significant as well. Thin films produced by deposition from solution have so far been moderately successful in terms of their mobilities^{11–13}, but the techniques are generally not suitable for high throughput. This hydrazine-based process can be applied more generally, and the hydrazinium salts need not be isolated first — they can be made *in situ*. If the process can be optimized and scaled up, thin films for high-performance channel layers in TFTs could be fabricated with all the processing performed at 300°C . In principle, depending on the specific metal chalcogenide involved, the films could be made at even lower temperatures.

However, the current processing temperature is too high for many applications (such as those using plastic substrates), and the mobilities achieved, although much higher than reported for other techniques, may not yet be adequate for many devices. Furthermore, the source and gate voltages of the TFTs are higher than those of typical silicon-based devices, while little is known about the yield and reproducibility of these devices. And the substrate is still silicon, not glass or plastic, which will limit the fabrication of TFTs on large-area, low-cost substrates.

So there are several factors to be considered before a new generation of optoelectronic devices based on this deposition

technology could gain a foothold, including the long-term operational and environmental stability of the devices. But, given the relative youth of this technology, and the exciting and rapid advances anticipated using chalcogenide thin films, the goal does not seem unattainable. Continued work in this area is likely to contribute to our understanding and exploitation of these exciting materials well into the next century.

Mercouri G. Kanatzidis is in the Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, USA.

e-mail: kanatzidis@chemistry.msu.edu

1. Mitzi, D. B., Kosbar, L. L., Murray, C. E., Copel, M. & Afzali, A. *Nature* **428**, 299–303 (2004).
2. Forrest, S. R. *Chem. Rev.* **97**, 1793–1896 (1997).

3. Sirringhaus, H. *et al. Science* **290**, 2123–2126 (2000).
4. Duan, X. *et al. Nature* **425**, 274–278 (2003).
5. Waldauf, C., Schilinsky, P., Perisutti, M., Hauch, J. & Brabec, C. J. *Adv. Mater.* **15**, 2084–2088 (2003).
6. Babel, A. & Jenekhe, S. A. *J. Am. Chem. Soc.* **125**, 13656–13657 (2003).
7. Meth, J. S., Zane, S. G., Sharp, K. G. & Agrawal, S. *Thin Solid Films* **444**, 227–234 (2003).
8. Kanatzidis, M. G. & Sutorik, A. C. *Prog. Inorg. Chem.* **43**, 151–265 (1995).
9. Enos, A. A. III, Liao, J.-H., Pikramenou, Z. & Kanatzidis, M. G. *Chem. Eur. J.* **2**, 656–666 (1996).
10. Dhingra, S. S. & Kanatzidis, M. G. *Mater. Res. Soc. Symp. Proc.* **180**, 825–831 (1990).
11. Gan, F. Y. & Shih, I. *IEEE Trans. Electron Devices* **49**, 15–18 (2002).
12. Yamaguchi, K., Yoshida, T., Sugiyama, T. & Minoura, H. *J. Phys. Chem. B* **102**, 9677–9686 (1998).
13. Sankpal, B. R., Mane, R. S. & Lokhande, C. D. *Mater. Res. Bull.* **35**, 177–184 (2000).

Cancer

Dangerous liaisons

Allan Balmain and Rosemary J. Akhurst

The cells of multicellular organisms are highly communicative and so can strongly influence one another's behaviour. One line of communication is particularly important in keeping cell growth in check.

A single cell destined to become a tissue or an organism can't go it alone in its rise to such dizzy heights. Communication, in the form of direct contacts between cells, interactions between cells and their surroundings, or the transmission of biochemical signals, is essential. Unravelling these networks of communication has provided gainful employment for biologists, geneticists and mathematicians in their quest to understand how the body forms¹. But now cancer biologists are being drawn into a similar web of interactions between cells targeted to become tumours (usually epithelial cells) and their neighbours (stromal fibroblasts). A network of signals operates in tumours. As they describe in *Science*, Bhowmick *et al.*² have identified one signalling pathway — regulated by transforming growth factor β (TGF- β) — that is an important mediator of the stromal–epithelial interactions modulating the growth of solid tumours.

It has been known³ for some years that normal stromal cells inhibit tumour growth whereas tumour-associated stromal cells stimulate it (Fig. 1). In their study, Bhowmick *et al.*² deleted the receptor for TGF- β — the TGF- β type II receptor — specifically in stromal cells of otherwise normal mice. This 'selective knockout' avoided killing the animals by deleting the TGF- β type II receptor in every cell type, completely inhibiting signalling through this pathway in the stroma of several tissues. Surprisingly, although the deletion occurs in the skin, oesophagus, kidney, liver and lung, mice were born normally, and these tissues showed no observable adverse effects.

Not everything, however, escaped unscathed. Prostate tissue underwent increased stromal-cell division, growing excessively by the time the animals were three weeks old. This, in turn, stimulated the epithelial cells of the prostate to divide and form lesions that resembled prostatic intra-epithelial neoplasia, a probable forerunner of prostate cancer. The stromal-cell population in the animals' forestomach also proliferated more rapidly, in this case spurring the expansion of the epithelial population so that an invasive form of cancer occurred that killed the mice by the time they were seven weeks old. So not only does abrogation of TGF- β signalling in the stromal fibroblasts cause them to proliferate, but the ensuing perturbed communication with the epithelial cells causes dysregulated cell division, indirectly leading to cancerous growth.

What causes this? Perhaps the stromal cells that cannot respond to TGF- β instead release other factors, or greater amounts of certain factors than do normal stromal cells? Bhowmick *et al.*² suggest that it might be due to another growth factor, hepatocyte growth factor (HGF), acting through its receptor c-Met (Fig. 1). The HGF–c-Met regulatory system is important in proliferation, cell migration and metastasis — the movement of cancer cells to other parts of the body to establish more tumours⁴. Impressively, fibroblasts from both the forestomach and prostate tissues of the knockout mice secreted at least three times as much HGF as their normal counterparts, and c-Met was simultaneously activated in the proliferating epithelial cells of the forestomach tumours.

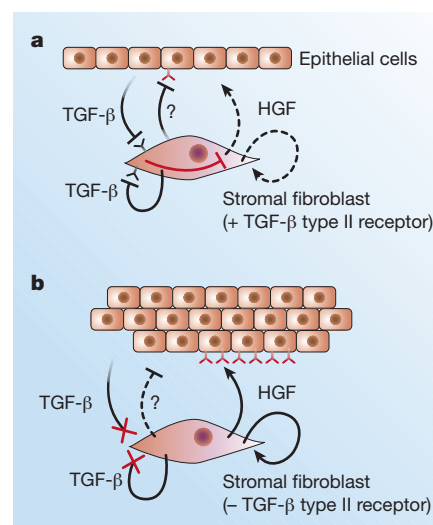


Figure 1 Cellular relationships. a, Normal communications between epithelial cells and their fibroblast neighbours. Both epithelial cells and fibroblasts secrete transforming growth factor β (TGF- β), which suppresses growth. Stromal fibroblasts might also secrete other factors that inhibit epithelial-cell growth (denoted by ?). A small amount of hepatocyte growth factor (HGF; its receptor is c-Met) secreted by the stroma inhibits stromal-cell growth and also that of epithelial cells. b, Perturbed signalling in the absence of the receptor for TGF- β , the TGF- β type II receptor. Inhibition of TGF- β signalling in stromal cells prevents growth-inhibitory responses to TGF- β and stimulates the stroma to release higher levels of HGF, a positive growth and metastatic factor. The production of other growth-inhibitory factors (?) might be reduced in response to inhibition of TGF- β signalling. TGF- β receptors are shown in black, c-Met receptors in red.

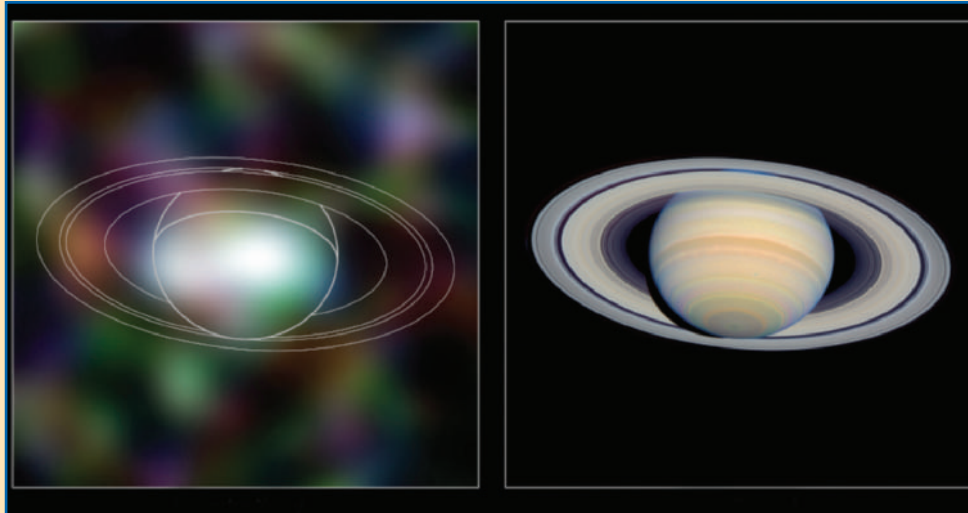
These data are consistent with previous reports that TGF- β normally inhibits HGF synthesis in stromal cells⁵. But they don't reflect the situation in advanced skin cancer, in which tumour-derived TGF- β induces adjacent stromal cells to produce HGF⁶.

The study by Bhowmick and colleagues has uncovered insights into cellular liaisons within tissues that should benefit cancer researchers and developmental biologists alike. But several issues raised by the findings must first be resolved. The cells of most solid tumours secrete large amounts of TGF- β , but are insensitive to its growth-inhibitory effects. This means either that components of this signalling pathway have mutated or, as is more common, that the growth response has been reduced while the ability to migrate, invade and metastasize in response to TGF- β is retained. How, then, do stromal cells normally escape the growth-inhibitory effects of overexpressed TGF- β and become willing partners in fostering epithelial tumour progression? Possible answers are genetic changes, or changes in gene expression that occur without altering the DNA sequence.

Planetary science

X-ray eyes on Saturn

NASA's Chandra X-ray Observatory has taken a long, clear look at the X-ray emission of Saturn — its unexpected findings are reported by J.-U. Ness and colleagues in *Astronomy & Astrophysics* (doi:10.1051/0004-6361:20035736; 2004). Although the rings and polar regions of the planet are 'silent', the equatorial region is beaming X-rays. Around 90 million watts of X-ray power is concentrated in the bright white spot in this image from Chandra (left, contrasted with an optical image of Saturn from the Hubble Space Telescope, right).



X-rays from other planets such as Mars and Venus are mostly only reflected rays from the Sun. But X-rays from Jupiter, Saturn's gas-giant neighbour, probably have an additional source. Earlier observations by Chandra have recorded strong X-ray emission from Jupiter's polar regions. This is believed to be a signature of auroral activity, caused by energetic heavy ions striking the jovian atmosphere and being directed by the planet's magnetic-field lines to the magnetic poles. At Earth, a similar light show is created by the energetic electrons of

the solar wind as they hit the atmosphere. Despite the different origins, the aurorae emit X-rays in both cases.

So what is happening on Saturn? The Hubble Space Telescope has provided striking images of Saturn's auroral activity at ultraviolet wavelengths, showing small haloes at the poles. One explanation for the absence of polar X-ray emission could be that the auroral phenomena on Saturn simply occur at lower energies. This would imply a different mechanism

for Saturn's aurorae, which is not entirely surprising — the presence of the rings could somehow alter the planet's magnetic-field lines.

The real puzzle is the brightness of the emission near Saturn's equator, compared with the darkness of its poles. If the Moon were scaled up to the size of Saturn and placed in its orbit, its X-ray brightness would still be about 50 times weaker than Saturn's. The emission from Saturn is, however, consistent with Jupiter's equatorial luminosity, so these gas giants are clearly much more

efficient at reflecting solar X-rays than is the Moon.

The stage is now set for Cassini, the NASA spacecraft on a dedicated mission to study the ringed planet. Cassini will enter Saturn's orbit on 1 July, and begin a four-year tour of the planet. And in December, the European Space Agency's Huygens probe, which has ridden piggyback on Cassini on the seven-year journey to Saturn, will separate from the NASA craft and enter the cloudy, nitrogen-rich atmosphere of Titan, Saturn's largest moon. **May Chiao**

It has generally been thought that any genetic changes occurring in the tumours are restricted to the epithelial cells, with stromal fibroblasts more often being portrayed as accessories rather than as fully active partners in tumorigenesis. This perception may change as more studies attempt to identify genetic changes in the stromal cells. In this context, two studies have demonstrated extensive loss of heterozygosity⁷ — deletion or disruption of one of two copies of a gene — or mutations in certain genes (*p53* or *PTEN*)⁸ in the stromal compartment of mammary tumours. Apart from the possibility that such populations of stromal cells may in fact represent 'real' tumour cells that have undergone a transition from epithelial cell to fibroblast⁹, this is clearly an area that warrants further investigation. Other data also show that loss or mutation of key genes in the host environment, rather than in tumour cells *per se*, can dramatically affect the behaviour of neighbouring cell populations^{10,11}.

An alternative to the stromal-mutation idea is the possibility that tumour-associated fibroblasts can act like fetal cells, acquiring an altered range of attributes, for example the ability to secrete a signal called migration-

stimulating factor^{12,13}. Interestingly, this signal is negatively regulated by TGF- β (ref. 13), raising the possibility that it, or a related molecule, may also play a role in producing the features observed by Bhowmick *et al.*².

Not surprisingly, the pharmaceutical industry is increasingly interested in developing inhibitors of TGF- β signalling to treat conditions ranging from fibrosis (excessive production of fibrous tissue) to advanced cancer. But might not such inhibitors promote epithelial tumours through the mechanism reported by Bhowmick *et al.*^{2,14}? Possibly, although Bhowmick and colleagues disrupted TGF- β signalling during embryonic development, rather than specifically in adult mice. Moreover, long-term inhibition of TGF- β signalling in mice, using soluble fragments of the TGF- β type II receptor that effectively sequester TGF- β but cannot process the signal, had no adverse effects — the mice were resistant to metastasis¹⁵.

An unexpected result from the study by Bhowmick *et al.*² is that some tissues, such as skin, in which TGF- β is known to be an important determinant of homeostasis, showed no obvious change in characteristics when TGF- β signalling was abrogated. So

the wiring of this signalling pathway varies in subtle ways between tissues, and gives us a foretaste of the surprises that may be in store as further experiments are carried out on signalling in specific subpopulations of tumour cells. ■

Allan Balmain and Rosemary J. Akhurst are at the Comprehensive Cancer Center, University of California, San Francisco, Box 0875, 2340 Sutter Street, San Francisco, California 94143, USA. e-mails: abalmain@cc.ucsf.edu rakhurst@cc.ucsf.edu

1. Gurdon, J. B. & Bourillot, P. Y. *Nature* **413**, 797–803 (2001).
2. Bhowmick, N. A. *et al.* *Science* **303**, 848–851 (2004).
3. Cunha, G. R., Hayward, S. W., Wang, Y. Z. & Ricke, W. A. *Int. J. Cancer* **107**, 1–10 (2003).
4. Trusolino, L. & Comoglio, P. M. *Nature Rev. Cancer* **2**, 289–300 (2002).
5. Sakurai, H. & Nigam, S. K. *Am. J. Physiol.* **272**, F139–F146 (1997).
6. Lewis, M. P. *et al.* *Br. J. Cancer* **90**, 822–832 (2004).
7. Moirand, F. *et al.* *Cancer Res.* **60**, 2562–2566 (2000).
8. Kurose, K. *et al.* *Nature Genet.* **32**, 355–357 (2002).
9. Derynck, R., Akhurst, R. J. & Balmain, A. *Nature Genet.* **29**, 117–129 (2001).
10. Bajenaru, M. L. *et al.* *Cancer Res.* **63**, 8573–8577 (2003).
11. Bajou, K. *et al.* *Nature Med.* **4**, 923–928 (1998).
12. Durning, P., Schor, S. L. & Sellwood, R. A. *Lancet* **2**, 890–892 (1984).
13. Ellis, I., Grey, A. M., Schor, A. M. & Schor, S. L. *J. Cell Sci.* **102**, 447–456 (1992).
14. Akhurst, R. J. *J. Clin. Invest.* **109**, 1533–1536 (2002).
15. Yang, Y. A. *et al.* *J. Clin. Invest.* **109**, 1607–1615 (2002).

Huntington's disease

Between genes and environment

Proc. Natl Acad. Sci. USA **101**, 3498–3503 (2004)

The fatal, inherited neurodegenerative disorder called Huntington's disease is caused by a three-base stutter in the DNA of the gene concerned. To a large extent, the length of the repeated section determines when the disease will strike. But there's the question of why people with identical repeats nonetheless succumb at different ages.

To answer this, Nancy S. Wexler and colleagues analysed the gene in 443 Venezuelans who suffer from the disease and are part of a much larger affected kindred. A statistical analysis confirmed that 72% of the variation in age of onset is caused by differences in the mutation itself.

Of the remaining variation in age of onset, 40% is determined by other 'modifier' genes that, for example, interact with the Huntington's disease gene or work in the brain cells that it destroys. The other 60% is determined by environmental influences such as diet, sanitation or exposure to pollutants.

One aim of continuing research is to try to identify the modifier genes and, potentially, drugs that might stall the onset or progression of the disease. Another is to pin down the most serious environmental risk factors, given the possibility that people who are susceptible to Huntington's disease might be able to minimize their exposure to such factors.

Helen Pearson

Planetary science

Marooned on Vesta

Earth Planet. Sci. Lett. **219**, 189–199 (2004)

Allan H. Treiman and colleagues have found veins of quartz in the Serra de Magé meteorite — evidence, they say, of ancient water on its parent asteroid Vesta.

Vesta lies in the main asteroid belt, between Mars and Jupiter. The rock of the

Serra de Magé meteorite is about 4.55 billion years old, but was changed by metamorphic activity about 4.4 billion years ago, while the meteorite was still part of its parent asteroid; the quartz veins formed at some time in between. The most likely explanation of their origin is that water carried free silica into cracks in the surrounding rock, where it was eventually transformed into quartz.

As there is no evidence that the water was originally part of Vesta, it must have arrived there sometime after the asteroid formed, but before the metamorphic activity began. Treiman *et al.* speculate that a comet might have delivered this water to Vesta, and that the asteroid may still carry polar ice deposits left by this collision.

Mark Peplow

Thermal physics

Balloon tricks

Phys. Rev. E (in the press); preprint at <http://arXiv.org/abs/cond-mat/0403171>

Most of us can predict what happens when two inflated rubber balloons — one large, one small — are connected by a hose: air will flow from the larger balloon into the smaller one until they are the same size. This seems obvious, and yet Yan Levin and Fernando L. da Silva have found that air can flow in the opposite direction.

The laws of thermodynamics tell us that a pressure difference between the balloons will drive air into the balloon with the lower pressure. But which balloon is which? This is determined by the surface tension of the balloons. In elastic objects, deformation reduces disorder (entropy), but nature prefers to maximize it. By calculating the surface tension of each balloon, the authors find that either balloon can have the higher pressure, depending on the combination of balloon-rubber elasticity and balloon size, so that air from the smaller balloon can indeed fill the larger one.

May Chiao

Plant physiology

How leaves stay dry

Langmuir doi:10.1021/la034961d (2004)

The waxy leaves of the lotus (*Nelumbo nucifera*) look smooth macroscopically, whereas plants such as *Alchemilla vulgaris*, also known as lady's mantle, have hairy leaves. Yet the surface of both remains dry after rain or morning dew.

It seems that plants have two distinct strategies for keeping their leaf surfaces dry, but the mechanisms involved are not fully understood. So physicists Alexander Otten and Stephan Herminghaus set out to do a more detailed study of Indian cress (*Tropaeolum majus*), which has smooth leaves like the lotus, and lady's mantle, using techniques such as scanning electron microscopy.

The surface of Indian cress leaves are



pitted with wax crystals, each a few micrometres long. This gives the leaf a roughness that traps pockets of air beneath the liquid, which the authors say is crucial for keeping the water droplets in tight spheres and makes them run off easily.

Conversely, the hairs on the leaves of lady's mantle are very hydrophilic. Small droplets form on the leaf's surface during dew condensation, but as soon as the hairs make contact with the water they clump together into bundles that stick into the droplet. The hydrophilic hairs might be expected to behave like blotting paper, but instead, their elasticity, coupled with the droplet's surface tension, creates a force that lifts the droplet away from the leaf's surface.

Mark Peplow

Tumour biology

Cut-off for blood supplies

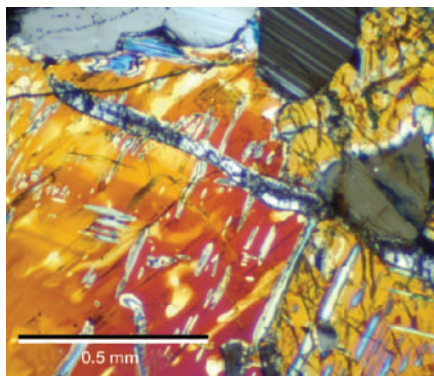
Cancer Res. **64**, 1570–1574 (2004)

Cyclophosphamide is an anticancer drug that can also shrink tumours by targeting their blood supply when given at lower but frequent doses. The molecule prompts endothelial cells — the cells that line blood vessels — to commit suicide, but its mechanism of action is unclear.

When tumour cells are treated with low doses of cyclophosphamide (LDC), the levels of a naturally occurring protein, thrombospondin-1, go up, report Yuki Hamano and colleagues. Thrombospondin-1 is anti-angiogenic — that is, it inhibits blood-vessel formation — so it may help to mediate the effects of LDC. Hamano *et al.* also show that LDC suppresses tumour growth in normal mice, but works less well in mice that lack thrombospondin-1. The effects seem to be specific to this protein, because mice lacking other natural anti-angiogenic agents, such as tumstatin and endostatin, respond well to LDC.

Given the implication that LDC will be most effective against tumours that produce thrombospondin-1, the authors say that assaying levels of thrombospondin-1 in biopsy samples could be used to help structure appropriate treatments.

Helen R. Pilcher



The Serra de Magé meteorite, seen here through crossed polarizers. A narrow, blueish vein of quartz runs from top left to lower right.



The cultural wealth of nations

Mark Pagel and Ruth Mace

Why, when the human race shows comparatively little genetic variation, are cultural differences so widespread and enduring? Thinking about cultures in terms of biological species provides some provocative answers.

Humans are the virtuosos of cultural diversity. We fish, hunt, shepherd, forage and cultivate. We practise polygyny, polyandry and monogamy, pay bride-prices and dowries, and have patrilineal and matrilineal wealth inheritance. We construct or inhabit all manner of shelters, speak about 7,000 different languages and eat everything from seeds to whales. And this is not counting many unique, and sometimes bizarre, belief systems and behavioural practices¹.

Contemporary cultural diversity is probably a fraction of the diversity that has ever existed, and there may be less cultural diversity now than there was 10,000 years ago, just before the advent of agriculture. Because of their superior reproductive rate, agriculturalists replaced many indigenous hunter-gatherer cultures as they spread across Europe and other parts of the globe^{2,3}.

If the picture of human cultures is one of variability, the human genetic landscape is one of homogeneity. All of humanity varies less genetically than does a typical wild population of chimpanzees^{4,5}. This may reflect our youthfulness as a species. Anatomically modern *Homo sapiens* emerged only about 75,000–100,000 years ago, and may have suffered a demographic 'bottleneck' in the recent past, meaning that in evolutionary terms we are all descended from a not-so-distant common ancestor. Also, of course, we can interbreed throughout our entire world-wide range. Add the facts that we regularly trade, migrate across each other's territories and wage war against each other, and a puzzle emerges: where does our extreme cultural diversity come from, and what maintains it?

The answers can perhaps be found in thinking about human cultures as if they are collections of distinct biological species. Just as species carry genetic adaptations to their environments, we believe that cultural adaptations have evolved in response to social life, and that such adaptations work to maintain cultural identity and coherence. Like species that do not interbreed, human cultures are surprisingly resistant to influences from other cultures and often act as barriers to gene flow. Important elements of culture show dominantly vertical modes of inheritance, much like the vertical transmission of genetic information. Collections of different species are observed to distribute themselves geographically in accordance with clear law-like patterns, and so, too, do human cultures. At a psychological level, humans display forms of social behaviour conducive to living in small groups, such as rewarding cooperation, punishing those who deviate from norms, and being wary of outsiders.

Cultures and species

But what is culture? This is one of those concepts that resists easy definition, even though most of us know which culture we belong to and who its other members are. We will use a simple definition that links cultures to distinct language groups (Box 1, overleaf), but it almost certainly underestimates their true diversity.

To see how human cultures partition the world like so many species, one need only look to linguistic diversity. Language–culture groups are not evenly distributed. Some 700 to 1,000 different languages, about 15% of the total on Earth, are spoken in the 312,000

square miles of the island of New Guinea⁶. In the northwest coastal areas a different language may be spoken every few miles. Similar densities of languages can be found on many Pacific island archipelagoes. By comparison, about 500 different languages were spoken in all of North America at the time of European contact, and only about 90 are spoken in China despite its vast population and area about 12 times the size of New Guinea. What determines variation in the density (numbers per unit area) of cultures?

One powerful factor is the environment. By 1953, the anthropologist J. B. Birdsall⁷ had documented the greater density of Australian Aboriginal tribal groups in wetter areas, and others have since reported higher cultural densities in coastal and equatorial regions⁸. In animals, a trend called Rapaport's rule⁹ holds that the density of animal species is highest in equatorial regions and declines steadily towards the poles. Polar regions support less biodiversity.

Some years ago, we tested Rapaport's rule in humans using information on Native American tribal groups at the time of European contact¹⁰. For each line of latitude, we recorded the number of languages spoken, and plotted its trend (Fig. 1, overleaf). The number of languages (human cultures) found per unit area is high in the lower latitudes and declines as one moves northwards. At extreme northern latitudes only a handful of different cultures inhabit the entire east–west expanse of northern Canada. An almost identical trend emerges in the diversity of different mammal species with latitude^{11,12}. In both cases the declines in diversity are matched by increasing sizes of

Box 1 Culture and language

A definition of culture has proved elusive, and yet this does not make the concept unusable. Biologists have similar difficulties defining species. The biological-species concept proposes that groups of animals form distinct species when they cannot interbreed successfully. Many apparently distinct species — such as horses and zebras or wolves and dogs — can interbreed. But what makes the concept

useful nonetheless is that these species normally do not interbreed if other options are available.

In this article we adopt a similarly pragmatic approach to cultures. *Collins English Dictionary*, in line with other sources (see refs 27, 28), defines a culture as “the sum total of a set of shared beliefs, values, practices”. Precisely what is shared and what is not will never be easy to document,

and, like species, not everyone will operate solely within their culture.

A rough guide to the number of different cultural groups can be obtained by taking them to be synonymous with distinct language groups, for which classification we rely on standard handbooks of the world's languages²⁹. This definition almost certainly underestimates the true number of distinct cultures. **M.P. & R.M.**

geographical ranges. Northerly dwelling animals are found over a large area, and so are northerly dwelling human cultural groups. A single species — humans — carved up North America as might 500 different species of mammal. Studies of African language-culture groups also find that biological and cultural diversity co-vary¹³.

Ecologists explain the trend in species diversity by noting that northern regions experience wider annual variation in climate. This favours species that are ecological generalists, able to cope with a range of conditions. The reverse is true for equatorial regions, where the relative annual stability selects for specialization in a particular habitat or niche. This explanation may be correct for animal species¹¹ but cannot work for humans. Apart from being just a single species, we are unequalled generalists, having colonized virtually the entire Earth through our ability to build fires, make clothes and shelters, and eat a variety of foods.

We believe that a more fundamental tendency accounts for the cultural patterns with latitude. The low cultural diversity seen in northerly latitudes is understandable, even expected. In this relatively unproductive environment, individuals must range over large areas to eke out a living. This movement of people tends to homogenize culture and language. The puzzle is to explain the higher cultural densities in more productive areas. Why do humans not just form one large and homogeneous cultural group in ecologically richer areas such as New Guinea?

It may be that human subpopulations continually split off from larger groups, and form around defensible resources — be they tracts of forest or fishing grounds — because individuals seek to control those resources. Similarities between human cultural and mammalian species diversity may then emerge because the same ecological factors make it more or less likely that small groups can be viable. Over time, these self-sufficient

groups emerge as daughter cultures and then as fully fledged different societies with their own customs, specializations and behavioural rules. The speed of cultural evolution makes this plausible. Even a new language can emerge after as few as 500 years of isolation. But we also suggest that it happens because human cultures have evolved mechanisms for maintaining their separation from other cultures.

Cultures and gene flow

For cultural variation to arise in the way we suggest, cultures must act to exclude each other. Anthropologists and linguists can construct family trees of major language groups based on the similarities between the languages. These trees — called phylogenies — chart the genealogical relationships between language groups. Since the late 1980s it has been known that phylogenetic trees of major language groups correspond closely to phylogenetic trees constructed from human genetic markers¹⁴. This result is consistent with our view, yet it may not be surprising: when, owing to

isolation for whatever reasons, people diverge from one another on background genetic markers, their languages and cultures also tend to drift apart.

But studies of European genetic diversity show that language differences may reduce genetic exchange between populations. Robert Sokal^{15,16} and colleagues measured the frequency of 63 genetic markers in samples taken from 3,119 different locations across Europe. They then applied the method of ‘wombling’ (after W. H. Womble¹⁷) to measure the rate of change in gene frequencies between these different locations. Standard genetic theory tells us that if people diffuse over an area such as Europe, then no boundaries of abrupt genetic difference are expected. However, barriers to the movement of people can produce zones of abrupt change.

Sokal *et al.* identified 33 boundaries separating areas of especially sharp changes in gene frequencies across Europe, and drew them on a map (Fig. 2). Twenty-two of the boundaries correspond to physical boundaries such as the Alps or the English Channel. They are also linguistic boundaries. For a further 11 boundaries, no physical or other barrier could be detected, and yet nine of these correspond to linguistic boundaries. In all, then, language is associated with regions of abrupt genetic change in 31 of the 33 cases. Not all studies support the separation of genes and language¹⁸, but it does seem that humans are more likely to mate with people they can talk to! The consequence is that cultural differences become self-reinforcing, prompting further cultural divergence.

Cultural transmission

Cultures also separate themselves from other cultures by the ways in which they transmit cultural information. Conventionally, we can describe vertical and horizontal modes of transmission of both genetic and cultural traits. The dominant mode of

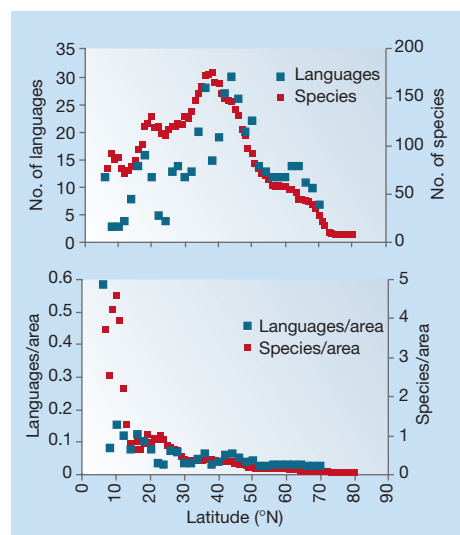


Figure 1 Language, mammals and Rapoport's rule. Top, Numbers of languages and of mammal species at each degree of north latitude in North America. The trends reflect the shape of the continent, being narrow in the south regions and growing wider at higher latitudes. Both trends peak at about 40° N, where North America is about 3,000 miles wide. (Data from refs 10 and 11.) Bottom, Densities of languages and mammal species, calculated as the number of either at a given latitude divided by the area of the continent for a 1° latitudinal slice at each latitude. Having controlled for area, both trends show the decline in density characteristic of Rapoport's rule⁹. Similar latitudinal trends for language and species diversity have been reported for Africa¹³.

transmission for both kinds of trait is vertical: parents transmit genes to offspring, and cultural traits are transmitted down generations by parents, elders and teachers. Nearly all of us speak our parents' language, and political and religious beliefs are surprisingly stable over generations. Humans are especially likely to copy common traits, a tendency that works to reduce variation within a culture but increase it between cultures. Genetic information is also sometimes transmitted horizontally, such as when a virus or other vector carries genetic material from one species to another. Normally, however, species barriers to interbreeding prevent large-scale horizontal transfer of genetic information, and it is an oddity when it is observed. We share more than 98% of our genes with chimpanzees because our genome is largely a closed shop to other genomes.

Barriers to horizontal transmission of culture are, in principle, far weaker. Horizontal transmission may sometimes be imposed, as when one culture conquers another. More prosaically, however, cultural traits can diffuse among geographically close neighbours, as when one culture borrows or adopts words, ideas, customs and technologies from another. The English language is an example, with about half of its vocabulary being of Germanic origin (vertical transmission) and half of Romance origin, reflecting, among other forces, the Norman conquest of England in the eleventh century. Of course, cultures do not always survive contact with one another, because a politically, technologically or economically dominant culture frequently displaces a weaker one.

It is revealing for our argument to consider how cultures treat events of horizontal transmission. Whereas vertical transmission of cultural traits goes largely unnoticed, horizontal transmission is far more likely to be regarded with suspicion or even indignation. France, for example, devotes a ministry to slowing or banning what is portrayed as an overwhelming march of English words, customs and phrases into the French language and culture. While the other nations of the world work away on their computers, the French sit at their *ordinateur*. The British are similarly alert to what they perceive to be 'Americanisms'. In both cases the home population is seen as an unwilling victim. However, a recent study of the English language found that it had admitted at least 90,000 new meanings over the past century, with only 5% derived from borrowings¹⁹. If there is an enemy, it is within. Cultures, it seems, like to shoot messengers.

These anecdotal accounts can be placed upon a firmer footing. If horizontal transmission is strong, cultures should tend to share their traits with those of their nearest geographical neighbours. If vertical transmission predominates, cultures

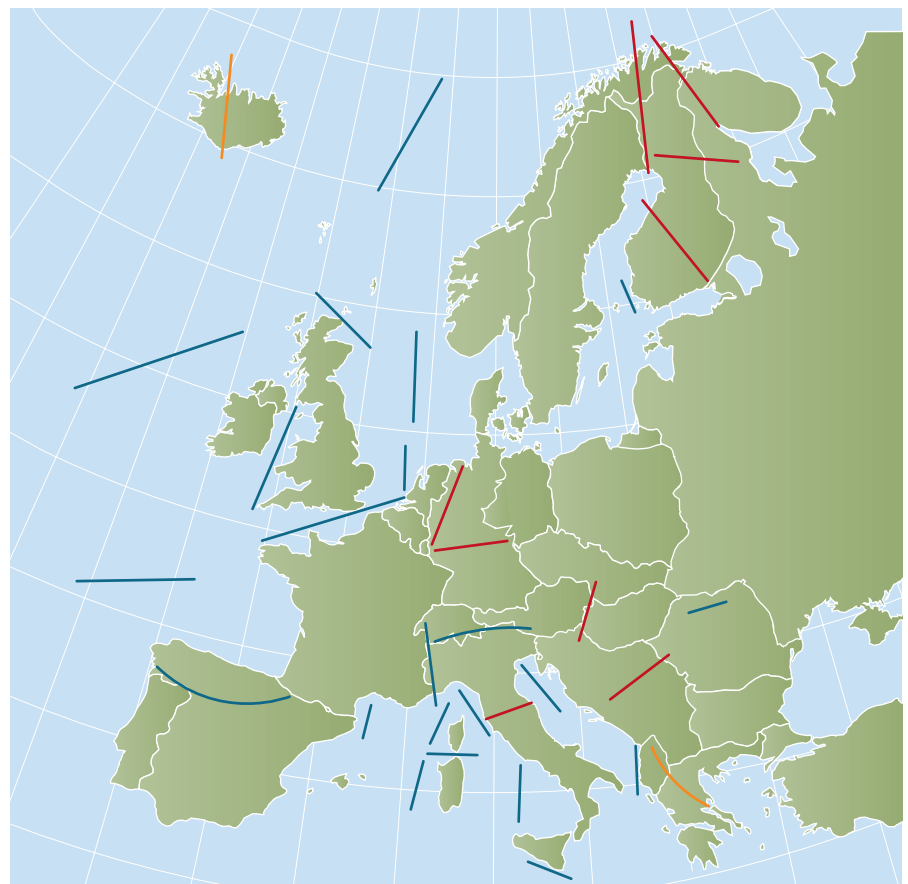


Figure 2 Genes and language in Europe. This map shows lines that divide regions of abrupt differences in the frequencies of 63 genetic markers in Europe's populations. The 22 blue lines identify regions that are separated by both physical barriers, such as the Alps or the English Channel, and linguistic differences. The 11 red and orange lines divide regions that have no detectable physical barriers, yet nine of the 11 (those in red) correspond to linguistic differences. In this case, then, language is associated with abrupt genetic change in 31 of 33 cases. Some lines fall in the sea (such as between Ireland and Iceland), indicating that the theoretical point of rapid genetic difference falls somewhere between the two locales and not on either one of them. (Map, redrawn from ref. 16, shows Europe of 1990, and provides a detailed key to the linguistic and other boundaries corresponding to the lines.)

should tend to have the traits of the cultures from which they descend; that is, of their ancestral culture (Box 2, overleaf). Although geography and ancestry are often correlated, a statistic known as the Mantel test²⁰ can distinguish the two hypotheses. In a study of 47 cultural traits in 277 African societies, most traits examined, especially those affecting means of subsistence, family structure and kinship (traits closely associated with reproductive success), were conserved over generations²¹. The only traits to show clustering that was dominantly geographical pertained to sexual division of labour. Another investigation, using a worldwide sample of cultures, found that subsistence practices, mating systems and sexual division of labour flowed down generations²². These results are striking: even under the influence of close geographical neighbours, cultures can remain stable and coherent units. Borrowing and imitation certainly occur. But cultural evolution is not a free-for-all in which all traits become equally available for adoption each generation.

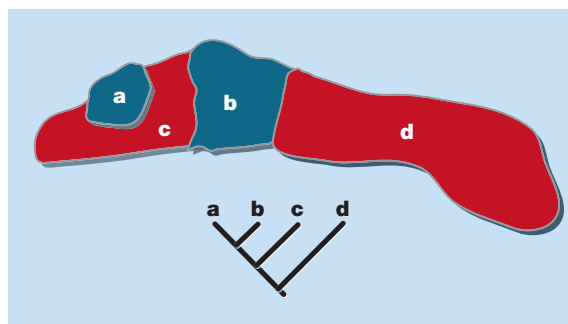
Extreme sociality

Human cultures are not as impermeable to outside influences as genomes, but they do erect barriers to the movement of people and ideas. What is so valuable as to call for this much protection? In a word, the answer is the group itself. It seems likely that human evolution in general, and human cultural evolution in particular, is distinguished by its sophisticated group behaviour. Historically, groups have had cooperative jobs to do, such as hunting large prey or warring with other human groups for access to territory or mates. These jobs are too large for any one person, placing a premium on group coherence, communication and cooperation.

The trouble is that this also makes cheating more profitable. The evolutionary theorist William Hamilton²³ argued that, to protect themselves, cooperative groups evolve strategies to make admission into their ranks difficult. These can take the form of being wary of outsiders, long periods of probation and costly (to the initiate)

Box 2 Transmission of cultural traits

The ancestor–descendant relationships between a group of cultures can be represented by a phylogenetic tree. Phylogenies are branching diagrams, like a genealogy or a family tree, that describe relatedness between a group of species or cultures. Linguistic data can be used to construct phylogenies of cultures for comparative cross-cultural studies^{30,31} in much the same way that genetic data can be used to construct phylogenies of animals. The diagram shows a hypothetical geographical area with four cultures labelled **a–d**. If geography determines the traits that any two cultures have in common, then, for example, cultures **a** and **c** should share the trait (the two



colours can be taken to represent two different values of some trait — for example, farming versus hunting and gathering as the culture's subsistence practice). If ancestry determines the value of the trait, then cultures that are near to one another on the phylogenetic tree should have the same value

of the trait, independently of geography. Studies show that ancestry (phylogeny) is often more influential than geography for core cultural traits^{21,22}. This is the outcome shown here: despite being next to culture **c**, culture **a** shares its trait with its nearest phylogenetic relative.

M.P. & R.M.

which individual over-exploitation of common resources results in their collapse, remind us of the price of selfishness. But this picture of the nature of cultures suggests that they are surprisingly robust against outside influences (although not invincible) and that, at least for large cultures, worries about cultural swamping are overstated. Nevertheless, our ancient cultural practices may also be telling us that, in a world in which mass movements of people from poorer to richer areas will become ever more common, we must be especially vigilant about our own tendencies to protect the status quo ante.

Mark Pagel is in the School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading RG6 6AJ, UK.

e-mail: m.pagel@reading.ac.uk

Ruth Mace is in the Department of Anthropology, University College London, Gower Street, London WC1E 6BT, UK.

e-mail: ucsadm@ucl.ac.uk

- Harris, M. *Cows, Pigs, Wars and Witches: The Riddles of Culture* (Vintage, New York, reissue edn, 1989).
- Renfrew, C. *Archaeology and Language: The Puzzle of Indo-European Origins* (Penguin, London, 1987).
- Diamond, J. & Bellwood, P. *Science* **300**, 597–603 (2003).
- Kaessmann, H., Wiebe, V. & Pääbo, S. *Science* **286**, 1159–1162 (1999).
- Kaessmann, H., Heissig, F., von Haeseler, A. & Pääbo, S. *Nature Genet.* **22**, 78 (1999).
- Wurm, S. in *Atlas of the World's Languages* (eds Moseley, C. & Asher, R. E.) 93–158 (Routledge, London, 1994).
- Birdsell, J. B. *Am. Nat.* **87**, 171–207 (1953).
- Nichols, J. *Evol. Anthropol.* **3**, 206–215 (1995).
- Rapaport, E. H. *Science* **201**, 679–686 (1982).
- Mace, R. & Pagel, M. *Proc. R. Soc. Lond. B* **261**, 117–121 (1995).
- Pagel, M. D., May, R. M. & Collie, A. *Am. Nat.* **137**, 791–815 (1991).
- Pagel, M. in *The Evolutionary Emergence of Language* (eds Knight, C., Studdert-Kennedy, M. & Hurford, J.) 391–416 (Cambridge Univ. Press, 2000).
- Moore, J. L. et al. *Proc. R. Soc. Lond. B* **269**, 1645–1653 (2002).
- Cavalli-Sforza, L. L., Piazza, A., Menozzi, P. & Mountain, J. *Proc. Natl Acad. Sci. USA* **85**, 6002–6006 (1988).
- Sokal, R. R. et al. *Am. Nat.* **135**, 157–175 (1990).
- Barbujani, G. & Sokal, R. R. *Proc. Natl Acad. Sci. USA* **87**, 1816–1819 (1990).
- Womble, W. H. *Science* **114**, 315–322 (1951).
- Rosser, Z. H. et al. *Am. J. Hum. Genet.* **67**, 1526–1543 (2000).
- Dent, S. *The Language Report* (Oxford Univ. Press, 2003).
- Mantel, N. *Cancer Res.* **27**, 209–220 (1967).
- Guglielmino, C. R., Viganotti, C., Hewlett, B. & Cavalli-Sforza, L. L. *Proc. Natl Acad. Sci. USA* **92**, 7585–7589 (1995).
- Holden, C. & Mace, R. *Am. J. Phys. Anthropol.* **10**, 27–45 (1999).
- Hamilton, W. H. in *ASA Studies 4: Biosocial Anthropology* (ed. Fox, R.) 133–153 (Malaby, London, 1975).
- Fehr, E. & Fischbacher, U. *Nature* **425**, 785–791 (2003).
- Gintis, H., Bowles, S., Boyd, R. & Fehr, E. *Evol. Hum. Behav.* **24**, 153–172 (2003).
- Kierkegaard, S. *Concluding Unscientific Postscript to Philosophical Fragments, 1846* (transl. Hong, H. V. & Hong, E. H.; Princeton Univ. Press, 1992).
- Durham, W. H. *Coevolution: Genes, Culture and Human Diversity* (Stanford Univ. Press, 1991).
- Cavalli-Sforza, L. L. & Feldman, M. W. *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton Univ. Press, 1981).
- Ruhlen, M. A. *A Guide to the World's Languages Vol. 1 Classification* (Arnold, London, 1991).
- Mace, R. & Pagel, M. *Curr. Anthropol.* **35**, 549–564 (1994).
- Pagel, M. in *Time-Depth in Historical Linguistics* (eds Renfrew, C., MacMahon, A. & Trask, L.) 189–207 (McDonald Inst. Archaeology, Cambridge, 2000).

Picture credits. Top row (left to right): F. Naumann/Panos; K. Mueller/Still Pictures; J. Miles/Panos; J.-L. Dugast/Still Pictures; E. Rassias/Gamma/Katz; J. Horner/Panos. Bottom row: Gamma/Katz; D. Tatlow/Panos; K. Robinson/Panos; J. Cancalosi/Still Pictures; C. & J. Lenars/Corbis; A. Wolfe/Still Pictures.

initiation ceremonies. This works to a degree, but cheats can still arise from within the group. Theorizing about group behaviour has therefore also emphasized the importance of genetic relatedness between group members. When it is high (as can be achieved if immigration is kept at a low level), altruistic and cooperative behaviours flourish because they tend to benefit one's own relatives. But genetic relatedness has its limitations. When it gets too high within a group, inbreeding begins to take its toll, and the benefits of sexual reproduction, namely the mixing and recombining of genes, are diminished.

Humans may have worked out a way to discard the need for relatedness as a means of ensuring cooperation: uniquely among the animals, humans may carry a set of behavioural adaptations specific to promoting cooperation and reciprocity, even when relatedness is low between group members^{24,25}. According to the doctrine known as strong reciprocity, humans are predisposed to cooperate with others, to make fair distribution of gains, and to punish those who fail to cooperate, even at a cost to themselves and with no expectation that these costs will be repaid^{24,25}.

This is a remarkable assertion, but should we believe it? Theoretical studies show that norms of cooperation and punishment of cheats can arise and be maintained in groups by a process of cultural 'group selection', in which more cohesive and cooperative groups outcompete groups riven by selfish cheats. Over time, these kinds of cooperative groups come to dominate. Laboratory experiments with volunteers and cross-cultural studies seem to support the strong-reciprocity view.

There are implicit norms of cooperation in groups, and individuals seem willing to punish cheats altruistically — that is, the punisher pays the cost of punishing — even though all group members reap the benefits.

This extreme sociality can make cooperation a stable strategy resistant to cheating even when group members are not related. But it does depend upon one key demographic feature: migration between groups must be kept low. If it is not, groups become homogenized, cheats can prosper and the driving force of group selection — differences between groups — fails. Wariness of strangers, for all its potentially ugly manifestations, may be deep in our psychological make-up.

An unscientific postscript

We end by taking our cue from the existential philosopher Søren Kierkegaard²⁶. If the view put forward here is correct, then human cultural diversity arises from two main forces. One is the drive to secede from larger groups whenever possible, the better to control some defensible resource; this is what gives rise to the geographical patterns of diversity. The second set of forces is social and behavioural. They maintain cooperation within groups and create cultural identity and coherence, causing barriers to gene flow and meaning that vertical cultural transmission dominates.

Putting these forces together, we get a picture of humans as a highly social and group-focused species. None of this is to say that selfish behaviour has been erased or that all cultures survive intact. The all-too-common 'tragedies of the commons', in

New mechanism for modulating colour vision

Single cones start making a different opsin as young salmon move to deeper waters.

Each cone photoreceptor in the retina responds to light in a limited range of wavelengths, giving it a spectral phenotype. This phenotype is determined by the most prevalent of the photoreceptor's visual-pigment proteins (opsins) and is assumed to remain unchanged during an animal's lifetime. Here we show that in the Pacific pink salmon, *Oncorhynchus gorbuscha*, single cones can switch their spectral phenotype from ultraviolet to blue by regulating the production of the appropriate opsins as the fish grow older. This photoreceptor plasticity may operate to modulate colour vision as the salmon's lifestyle changes.

The vertebrate retina has several spectral types of cone photoreceptor. Each photoreceptor contains a visual pigment that is maximally sensitive to ultraviolet (UV), blue, green or red light¹ and consists of a protein (opsin) attached to a chromophore (a derivative of vitamin A₁ or vitamin A₂)¹. It is the combined input from cones containing different visual pigments, each with a different spectral phenotype, that allows an animal to perceive colour.

Changes in colour sensitivity in the outer retina arise from altering the density of different cone types² or by switching chromophores (vitamin A₂-based visual pigments absorb light of longer wavelengths than their vitamin A₁ counterparts)³. Another way to modulate colour vision would be to switch between different opsins, a mechanism that we investigate here in the cone photoreceptors of the Pacific pink salmon.

To determine whether it is possible for such a phenotypic transformation to occur: we measured the absorbance of visual pigment along the outer segment of individual cones by using microspectrophotometry³ (the outer segment of a cone consists of stacked lipid bilayers that contain the visual pigment)¹; we also labelled opsin messenger RNAs by using *in situ* hybridization⁴ with salmon-specific molecular probes.

We found that all the single cones in recently hatched fish (weight about 0.4 g) contain a visual pigment with maximum absorbance in the ultraviolet ($\lambda_{\max}$ of the visual pigment: 365–375 nm); however, as the fish grew (weight exceeding 0.8 g), the single cones stopped producing UV-opsin and started to produce blue-opsin ($\lambda_{\max}$ of the visual pigment: 425–435 nm) in a transformation event that proceeds from the ventral to the dorsal retina.

When absorbance is measured from the tip to the base of a cone outer segment undergoing such a transformation, the absorbance is ultraviolet-dominated at the

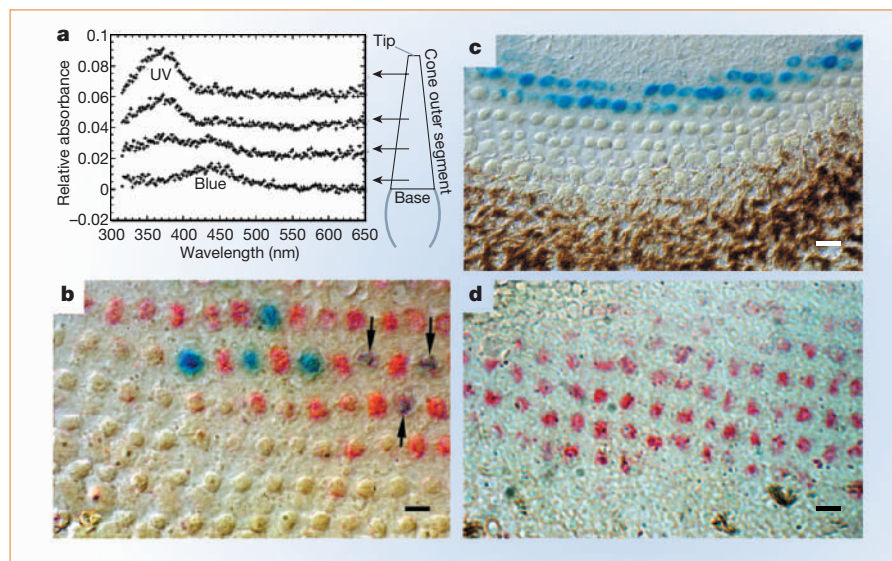


Figure 1 Absorbance measurements and double *in situ* labelling of single cones in Pacific pink salmon with ultraviolet- and blue-opsin riboprobes. **a**, Cones show increased absorbance in the ultraviolet at the tip of the outer segment (right) and increased absorbance of blue light at the base, with intermediate absorbances in between (traces are composite averages from seven cones). **b**, Tangential retinal section showing single cones labelled by *in situ* hybridization with probes specific for mRNAs encoding UV-opsin (shown in blue) or blue-opsin (in red); cones coexpressing both mRNAs appear purple (arrows). **c,d**, Single cones are labelled with only the UV-opsin riboprobe before the switch in opsin expression (**c**), and with only the blue-opsin riboprobe afterwards (**d**). Scale bars: **b**, 5.5 μ m; **c**, 7.8 μ m; and **d**, 13 μ m.

tip, becoming blue-dominated at the base (Fig. 1a). As the outer-segment bilayers are produced at the base and removed from the tip⁵, these results show that the UV-opsin is being replaced by blue-opsin during the transformation.

In agreement with this finding, fish in the process of transformation have a mixture of ultraviolet- and blue-sensitive cones, with some cones expressing both opsins together (Fig. 1b). Before the transformation, single cones express only UV-opsin (Fig. 1c); afterwards, they express only blue-opsin (Fig. 1d). Salmonid fish also lose some single (corner) cones following this transformation², further reducing their ultraviolet/blue sensitivity.

This opsin switch in the single cones of pink salmon represents a previously undiscovered way to modulate colour vision. The transformation is linked to a progressive change in the lifestyle of the salmon⁶, which starts as a planktivore in surface waters, where ultraviolet light is abundant, and becomes a fish-eating predator in deeper waters, where blue-green light prevails. As several other vertebrates are known to co-express different opsins in a cone^{7–10}, such a mechanism for temporal modulation of colour vision may be widespread.

Christiana L. Cheng,

Iñigo Novales Flamarique

Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

e-mail: inigo@sfu.ca

1. Fein, A. & Szuts, E. Z. *Photoreceptors: Their Role in Vision* (Cambridge University Press, Cambridge, UK, 1982).
2. Kunz, Y. W., Wildenburg, G., Goodrich, L. & Callaghan, E. *Vision Res.* **34**, 1375–1383 (1994).
3. Hárosi, F. I. *Vision Res.* **34**, 1359–1367 (1994).
4. Forsell, J., Ekström, P., Novales Flamarique, I. & Holmqvist, B. *J. Exp. Biol.* **204**, 2517–2525 (2001).
5. Besharse, J. C. in *The Retina: A Model for Cell Biological Studies* Part 1 (eds Aler, R. & Farber, D.) 297–352 (Academic, New York, 1986).
6. Groot, C. & Margolis, L. *Pacific Salmon Life Histories* (UBC Press, Vancouver, 1991).
7. Makino, C. L. & Dodd, R. L. *J. Gen. Physiol.* **108**, 27–34 (1996).
8. Szél, Á., Van Veen, T. & Röhlich, P. *Nature* **370**, 336 (1994).
9. Lukáts, Á. *et al. Invest. Ophthalmol. Vis. Sci.* **43**, 2468–2473 (2002).
10. Parry, J. W. L. & Bowmaker, J. K. *Invest. Ophthalmol. Vis. Sci.* **43**, 1662–1665 (2002).

Competing financial interests: declared none.

Buckminsterfullerenes

A non-metal system for nitrogen fixation

In all nitrogen-fixation processes known so far — including the industrial Haber–Bosch process¹, biological fixation by nitrogenase enzymes² and previously described homogeneous synthetic systems^{3–5} — the direct transformation of the stable, inert dinitrogen molecule (N₂) into ammonia (NH₃) relies on the powerful redox properties of metals. Here we show that nitrogen fixation can also be achieved by using a non-metallic buckminsterfullerene (C₆₀) molecule, in the

form of a water-soluble C_{60} : γ -cyclodextrin (1:2) complex⁶, and light under nitrogen at atmospheric pressure. This metal-free system efficiently fixes nitrogen under mild conditions by making use of the redox properties of the fullerene derivative.

Treatment of a γ -cyclodextrin-bicapped C_{60} complex⁶ (**1** in Fig. 1) with 100 equivalents of sodium bisulphite ($Na_2S_2O_4$) as a reducing reagent in water under 1 atmosphere of nitrogen and with visible light at 60 °C for 1 hour gave ammonia in 33% yield (based on the amount of C_{60} in the reaction); prolonging the reaction time to 24 hours did not increase the yield. The combination of $Na_2S_2O_4$, **1** and N_2 is essential for the reaction, as control reactions lacking any one of these reactants produced no ammonia. Replacing **1** by γ -cyclodextrin, C_{60} or a simple mixture of C_{60} and γ -cyclodextrin, or performing the reaction at 25 °C, failed to produce ammonia.

The pH of the reaction must be kept neutral, as yields of ammonia were lower under acidic or basic conditions. When the reaction was carried out under an atmospheric pressure of $^{15}N_2$, the formation of $^{15}NH_4^+$ ($^{15}ND_4^+$; -363 p.p.m. in D_2O) was confirmed by ^{15}N NMR spectroscopy after acidification of the crude product. This result confirms that molecular nitrogen undergoes conversion into ammonia in this system.

Visible light was also essential to the reaction as no ammonia was formed in the absence of light. An increased yield of ammonia (45%) was obtained when the reaction mixture was irradiated with a high-pressure mercury lamp (450 W) without a slit. Strong absorption in the ultraviolet region, with weaker but significant bands in the visible region, was seen in the spectrum of **1** in water: characteristic features of the C_{60} moiety were observed at wavelengths of 212 (strong), 259 (strong), 333 (strong), 408 (medium), 536 (weak) and 601 (weak) nanometres (results not shown).

These results indicate that charge transfer between nitrogen and the reduced and electronically excited **1** (exciplex formation) is probably involved in the reduction of N_2 , given that C_{60} and its derivatives are easily excited by low-energy light and that their excited states readily promote electron and energy transfer compared with the ground-state species⁷. Light may therefore assist electron and energy transfer from C_{60} in **1** to N_2 (probably coordinated to **1**). A similar effect of light has been described for a transition-metal complex⁸.

We consider that all of the electrons required for the formation of ammonia are supplied by the reduced C_{60} in **1**, because C_{60} in the complex is readily reduced to the corresponding C_{60}^- and C_{60}^{2-} species by using $Na_2S_2O_4$ (ref. 9). The cyclic voltammogram of **1** (in water under both argon and nitrogen atmospheres) reveals two waves: a reversible

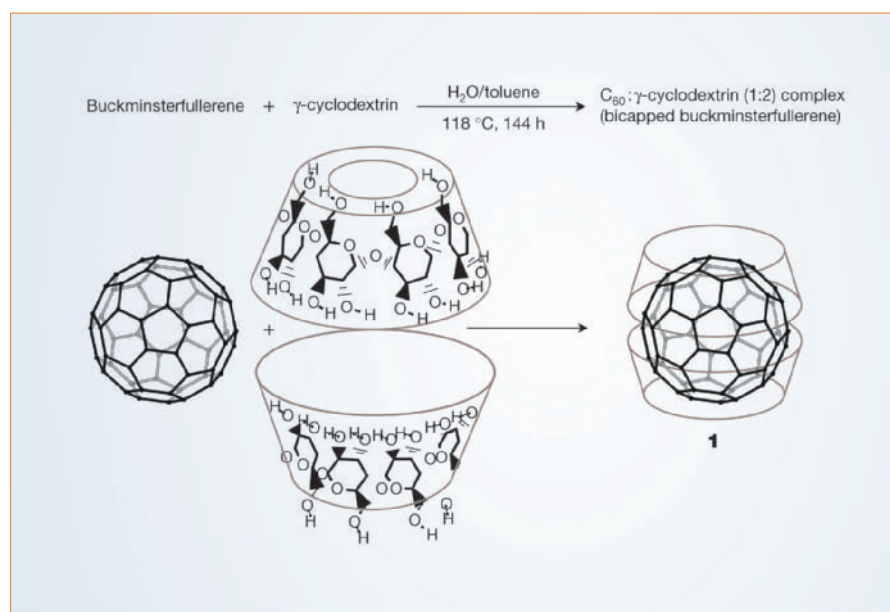


Figure 1 Preparation of γ -cyclodextrin-bicapped C_{60} complex (**1**), which is used in the fixation of nitrogen to ammonia. A mixture of C_{60} (0.400 g; 0.555 mmol) and γ -cyclodextrin (1.200 g; 0.925 mmol) was stirred in a water (160 ml)/toluene (60 ml) mixture at 118 °C for 48 h and then γ -cyclodextrin (0.600 g; 0.463 mmol) was added twice more every 48 h; complex **1** was produced in 70% yield (1.464 g; 0.391 mmol) as a purple solid, with each molecule being coordinated with 24 H_2O molecules (stoichiometry from thermogravimetric and differential thermal analysis). In the fixation reaction, a suspension of **1** (37.4 mg; 0.010 mmol) under 1 atmosphere of nitrogen with $Na_2S_2O_4$ (174 mg; 1.00 mmol) in water (10 ml) was magnetically stirred at 60 °C for 1 h under visible light from a fluorescent lamp. The yield of ammonia was quantified by using indophenol reagent.

wave at -0.61 V (the half-wave electric potential) assignable to the one-electron redox process C_{60}/C_{60}^- ; and an irreversible wave at -1.09 V (the cathodic peak potential) assignable to the one-electron redox process C_{60}^-/C_{60}^{2-} . A saturated calomel electrode was used as the reference electrode. The results indicate that the reduced complexes of **1** have sufficiently low reduction potentials⁵ to reduce the coordinated N_2 molecule.

Some molybdenum and tungsten complexes with nitrogen can be protonated by sulphuric acid to produce ammonia in high yields as a result of the presence of the metal atoms³. We therefore investigated whether protonation occurs in C_{60} -coordinated N_2 . The reaction of **1** with an excess of dilute sulphuric acid under 1 atmosphere of nitrogen at 60 °C for 1 hour produced ammonia in 15% yield (based on C_{60}). Control experiments run for over 1 hour gave no ammonia at 25 °C, or when **1**, H_2SO_4 or N_2 were absent. The yield of ammonia was lower than that obtained in the reaction using $Na_2S_2O_4$. Visible light was again essential for reaction, with the yield of ammonia being increased to 25% under a high-pressure mercury lamp (as specified above).

The infrared and Raman spectra of **1** unfortunately showed no clear-cut absorption that could be assigned to the $\nu_{N=N}$ band of coordinated N_2 on **1**. No peaks other than that due to $^{15}N_2$ (-75 p.p.m. in D_2O) were observed by ^{15}N NMR on treatment of **1** in D_2O under one atmospheric pressure of $^{15}N_2$. These results indicate that complex formation between N_2 and **1** is not

sufficient to be detected by infrared, Raman and NMR spectroscopy. However, it is possible that molecular N_2 might coordinate to the surface of the C_{60} molecule in **1** in an end-on mode.

Yoshiaki Nishibayashi*, **Makoto Saito***, **Sakae Uemura***, **Shin-ichi Takekuma†**, **Hideko Takekuma†**, **Zen-ichi Yoshida†**

*Department of Energy and Hydrocarbon Chemistry, Graduate School of Engineering, Kyoto University, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan
e-mail: uemura@scl.kyoto-u.ac.jp

†Department of Applied Chemistry, Faculty of Science and Engineering, Kinki University, 3-4-1 Kowakae, Higashi Osaka, Osaka 577-8502, Japan

1. Appl. M. Ammonia (Wiley-VCH, Weinheim, 1999).
2. Postgate, J. *Nitrogen Fixation* (Cambridge University Press, Cambridge, 1998).
3. Chatt, J., Pearson, A. J. & Richards, R. L. *Nature* **253**, 39-40 (1975).
4. Nishibayashi, Y., Iwai, S. & Hidai, M. *Science* **279**, 540-542 (1998).
5. Yandulov, D. V. & Schrock, R. R. *Science* **301**, 76-78 (2003).
6. Yoshida, Z., Takekuma, H., Takekuma, S. & Matsubara, Y. *Angew. Chem. Int. Edn* **33**, 1597-1599 (1994).
7. Guldi, D. M. & Prato, M. *Acc. Chem. Res.* **33**, 695-703 (2000).
8. Solari, E. et al. *Angew. Chem. Int. Edn* **40**, 3907-3909 (2001).
9. Takekuma, S., Takekuma, H., Matsumoto, T. & Yoshida, Z. *Tetrahedron Lett.* **41**, 2929-2932 (2000).

Competing financial interests: declared none.

brieF communications arising online

www.nature.com/bca

Earth science: An alternative origin for the 'Silverpit crater'

J. R. Underhill (doi:10.1038/nature02476)

Reply: S. A. Stewart & P. J. Allen

(doi:10.1038/nature02480)

Earth science

An alternative origin for the 'Silverpit crater'

Arising from: Stewart, S. A. & Allen, P. J. *Nature* **418**, 520–523 (2002).

Recent interpretation of three-dimensional seismic data in the southern North Sea has led to the recognition of a concentric, multi-ringed depression (the 'Silverpit crater'), the genesis of which Stewart and Allen attribute to a meteor strike¹. Although this interpretation has understandably excited the scientific community, an independent evaluation of the well calibrated and densely spaced seismic data indicates that the structure may have a different origin. This new interpretation attributes its formation to salt withdrawal at depth, rather than to extra-terrestrial impact.

Seismic lines across the Silverpit crater demonstrate the folded nature of the sedimentary section, which affects all stratigraphic levels down to and including the top Permian (Fig. 1). Each fold has a structural relief of 1-second, two-way travel time (that is, over 2 km). The Silverpit crater lies in the core of one of the two prominent synclines lying between three antiforms. As the top Permian and basal Permian (top Rotliegend Group) seismic reflectors are not parallel to each other, the intervening Zechstein Group clearly varies in thickness across the area. Significantly, the synclinal configuration of the top Permian and younger stratigraphic horizons and the thinning of the Zechstein Group are exactly coincident, a fact that was not mentioned by Stewart and Allen.

Well observations in the Silverpit area indicate that the Zechstein Group predominantly consists of highly deformed evaporite deposits. Their burial promoted mobility on geological timescales, leading to complete withdrawal and grounding of the sedimentary overburden in some areas². Conversely, influx and buoyant rise of evaporites has led to the formation of salt pillows and diapirs that fold and penetrate the overburden elsewhere².

Given the local structure, scale (both amplitude and wavelength) and regional considerations, the obvious conclusion is that the structural depression previously attributed to meteor impact has an unusual but more mundane origin as one of a number of depressions created by the almost complete withdrawal of support at depth. Stratigraphic onlap of seismic reflectors within the sedimentary succession demonstrates that the last phases of its growth occurred during Cenozoic deposition. Truncation beneath the base Cretaceous seismic reflector suggests that an earlier phase of evaporite mobility also occurred.

Two other aspects of the original interpretation¹ are also worthy of discussion as both were interpreted to support bolide impact. These are the development of

annular fracture patterns and the apparent occurrence of a central conical peak.

The concentric array of normal faults affecting the top Cretaceous horizon is attributed here to deformation that accommodated folding². Analogous ring faults occur in association with other salt withdrawal basins³. Other examples of extensional deformation in such settings include fracture patterns related to caldera collapse following volcanic eruption⁴, ice-melt collapse pits⁵, ice cauldrons on the top

of ice sheets resulting from subglacial heating⁶ and mining subsidence. Given the plausible alternatives, comparison with Valhalla-type multi-ring basins should not be used as vindication of meteor impact.

Evidence for the presence of the central peak is not unequivocal. Indeed, the seismic data show that all of the post-Permian stratigraphic horizons (for example, the top Cretaceous) reach their greatest depths in the core of the syncline. The difficulty in substantiating the peak's existence leaves open the possibility that it was an artefact of the seismic processing, velocity modelling or the method of interpretation itself.

In conclusion, although publication of the image of the Silverpit crater demon-

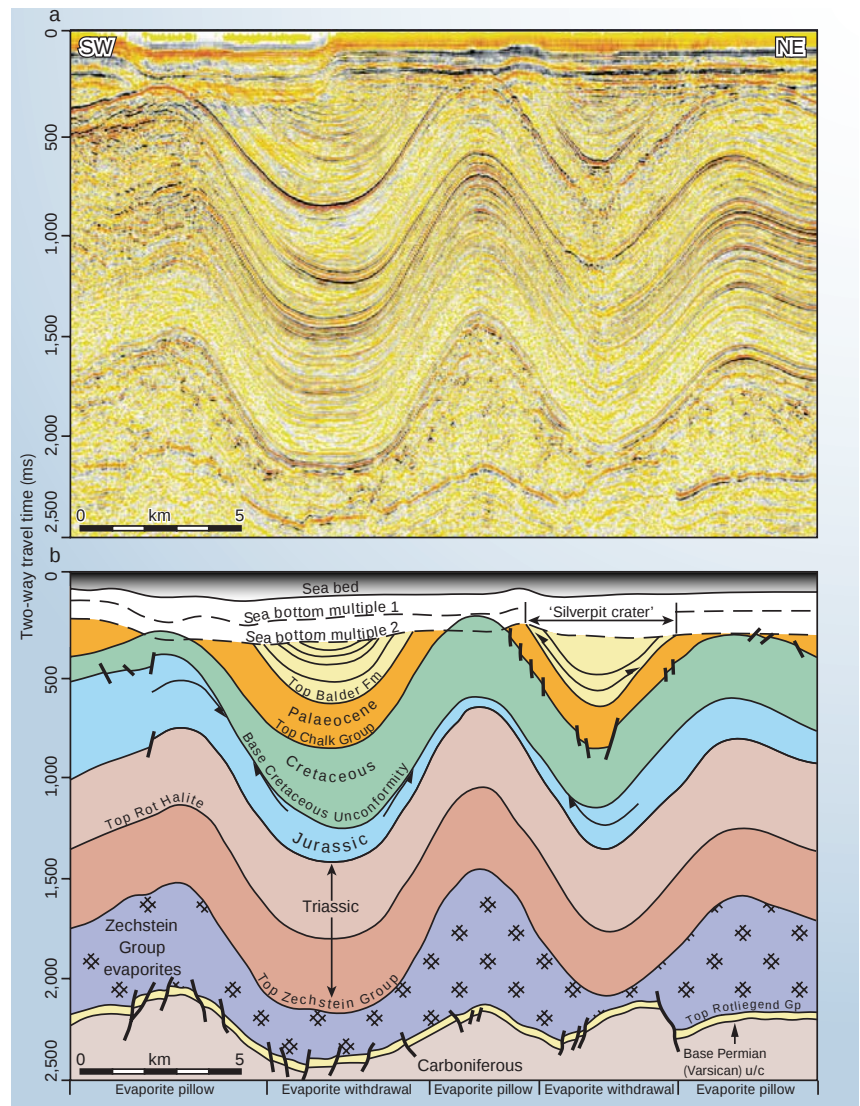


Figure 1 Seismic line and corresponding geological cross-section through the Silverpit area. **a**, The line transects the centre of the Silverpit crater and demonstrates the structural relief of folding and the exact coincidence between the synclines and thinning of Upper Permian Zechstein Group evaporites at depth. The implication is that the feature is simply one of a number of unusual but similar structures that resulted from salt withdrawal at depth, rather than from meteor impact. **b**, Stratigraphic relationships imply that the structure did not result from a single episode of evaporite mobility but had a long-lived growth history that started in the Mesozoic and continued in the Cenozoic. The section is vertically enlarged by about three times for clarity.

strates the added value provided by the interpretation of seismic volumes, ascribing its genesis to meteor impact remains speculation. In the absence of independent supporting data, an origin through salt withdrawal is not only more plausible but also consistent with the geological history of the basin.

John R. Underhill

*Grant Institute of Earth Science, School of
Geosciences, University of Edinburgh, King's
Buildings, West Mains Road,
Edinburgh EH9 3JW, UK
e-mail: jru@staffmail.ed.ac.uk
doi: 10.1038/nature02476*

1. Stewart, S. A. & Allen, P. J. *Nature* **418**, 520–523 (2002).
2. Jenyon, M. K. *J. Geol. Soc. Lond.* **145**, 445–454 (1988).
3. Maione, S. J. *The Leading Edge* **20**, 818–829 (2001).
4. Branney, M. J. *Bull. Volcanol.* **57**, 303–318 (1995).
5. Branney, M. J. & Gilbert, J. S. *Bull. Volcanol.* **57**, 293–302 (1995).
6. Wood, C. A. *Proc. Lunar. Planet. Sci.* **12a**, 173–180 (1981).

Stewart and Allen reply — The coincidence of the Silverpit crater and a Tertiary fold axis is curious. There can be no doubt that the regional folds are detachment structures accommodated by flow of Permian evapor-

ites. It is quite straightforward, however, to demonstrate that the crater itself is not related to salt withdrawal as Underhill¹ proposes.

The Triassic strata, which separate the Silverpit structure from the Permian evaporites and contain minor evaporite layers, do not display any structures of a comparable wavelength or form to the Silverpit crater, which, as we and Underhill agree, is a sharp-edged structure in the top chalk². In other words, there is no evidence for the Silverpit crater having formed by propagation of roof collapse from sub-Cretaceous level.

Unfolding or 'flattening' any of the key stratigraphic markers can further emphasize this, as previously approximated by presentation of a seismic section parallel to the fold axis² (a geometrically permissible approach due to radial symmetry of the crater). With the absence of short-wavelength structure in the stratigraphic markers in between the crater and the salt, there cannot be structural association between salt movement and crater formation. Indeed, all cross-sections including Underhill's Fig. 1b show that the crater is accommodated by volume loss in the Cretaceous section alone².

It is also worth noting that the main fold event is significantly younger than the crater, as recorded by a comparable degree of folding of the top Palaeocene and top Cretaceous (see ref. 2 and Underhill's Fig. 1b). Thus the crater was already fully formed, and buried, before the main episode of folding. Still, the puzzle of why the fold axis intersects the crater remains. We suggest the opposite sense of causality to that offered by Underhill. It is well known that flaws and weaknesses serve as nuclei and attractors for structural failure³. The Silverpit crater may have performed this role at a kilometre scale and been 'first to fail' during post-crater regional shortening, thereafter associated with the axis of an amplifying fold.

Simon A. Stewart*, Philip J. Allen†

**Chirag Well Planning, BP Azerbaijan,
c/o Chertsey Road, Middlesex TW16 7LN, UK
e-mail: stewarsa1@bp.com*

*†Production Geoscience Ltd, Banchory
AB31 5YR, UK*

doi: 10.1038/nature02480

1. Underhill, J. R. *Nature* doi:10.1038/nature02476 (2004).
2. Stewart, S. A. & Allen, P. J. *Nature* **418**, 520–523 (2002).
3. Kuksenko, V. et al. *Pure Appl. Geophys.* **146**, 253–263 (1996).

Control of gene expression by a natural metabolite-responsive ribozyme

Wade C. Winkler^{1*}, Ali Nahvi^{2*}, Adam Roth^{1*}, Jennifer A. Collins^{1*} & Ronald R. Breaker¹

¹Department of Molecular, Cellular and Developmental Biology, and ²Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 208103, New Haven, Connecticut 06520-8103, USA

* These authors contributed equally to this work

Most biological catalysts are made of protein; however, eight classes of natural ribozymes have been discovered that catalyse fundamental biochemical reactions. The central functions of ribozymes in modern organisms support the hypothesis that life passed through an 'RNA world' before the emergence of proteins and DNA. We have identified a new class of ribozymes that cleaves the messenger RNA of the *glmS* gene in Gram-positive bacteria. The ribozyme is activated by glucosamine-6-phosphate (GlcN6P), which is the metabolic product of the GlmS enzyme. Additional data indicate that the ribozyme serves as a metabolite-responsive genetic switch that represses the *glmS* gene in response to rising GlcN6P concentrations. These findings demonstrate that ribozyme switches may have functioned as metabolite sensors in primitive organisms, and further suggest that modern cells retain some of these ancient genetic control systems.

Natural ribozymes catalyse fundamental reactions required for RNA processing^{1–4} and protein synthesis⁵ that are essential to all cellular life. In addition, numerous functional RNAs have been created in the laboratory by using directed evolution^{6,7}, suggesting that nucleic acids have substantial untapped potential for sophisticated function. Presumably, such functional RNAs could take on even more diverse biological roles such as binding small metabolites and regulating key cellular processes. Consistent with this hypothesis is the fact that allosteric ribozymes that respond to specific target molecules can be created using molecular engineering strategies^{8–10}, and they retain their function even in complex biochemical mixtures¹¹. These findings strongly suggest that similar metabolite-responsive ribozymes might have emerged through natural evolutionary processes to serve as genetic switches.

The control of gene expression in modern cells seems to be dominated by protein factors¹². However, it has been determined recently that specialized domains within certain mRNAs selectively bind metabolites and control gene expression without the need for protein factors^{13–16}. In prokaryotes, these 'riboswitches' harness allosteric changes that result when metabolites are bound by RNA receptors (aptamers) to control either transcription termination or translation initiation¹⁷. For example, metabolite binding by purine-specific riboswitches¹⁸ guides the differential formation of either an intrinsic terminator stem^{19,20} or an anti-terminator stem to control the synthesis of mRNAs. Variations of this relatively simple mechanism of alternative folding are known to form genetic 'on' or 'off' switches as needed by the cell^{18,21}. Even greater diversity of genetic control mechanisms would be possible if the action of natural ribozymes were to be controlled by metabolite binding.

As riboswitches do not require protein factors for function, and as they are used to sense compounds that are fundamental to nearly all living systems, we have speculated^{13,14} that organisms from the RNA world might have maintained a sophisticated metabolic state using similar allosteric RNAs. This hypothesis is supported by repeated observations that riboswitches are broadly distributed among many distantly related organisms^{17,22} including eukaryotes^{23,24}, which suggest that these RNAs emerged early in evolution. Herein, we report the discovery of a genetic control element based on a metabolite-responsive ribozyme. This finding highlights the possible diversity of ancient RNA functions, and suggests that sophisticated genetic control systems in modern organisms can be dependent on metabolite-triggered ribozymes.

A novel RNA element

Recently, we have identified a number of new sequence elements residing in intergenic regions of the *Bacillus subtilis* genome that produce highly structured RNA domains when prepared by transcription *in vitro*²⁵. The sequences and structures of these domains compare favourably with those of known riboswitches¹⁷, which are typically larger and more elaborately folded than RNA motifs that are recognized by protein genetic factors. Furthermore, most of these elements are widespread among bacterial lineages, which is consistent with the distributions of most known riboswitches^{22,23}. Therefore, we have begun to investigate whether these new RNA elements bind metabolites and control gene expression.

Most intriguing to us was the conserved element that resides upstream from the *glmS* gene in *B. subtilis* and at least 17 other Gram-positive bacteria. The putative 5' untranslated region (UTR) of each *glmS* mRNA in the phylogeny²⁵ exhibits a high level of sequence conservation and numerous instances of nucleotide covariation that are consistent with the formation of several secondary-structure elements (Fig. 1a, P1–P4). The *glmS* gene encodes the enzyme glutamine-fructose-6-phosphate amidotransferase, which uses fructose-6-phosphate (Fru6P) and glutamine to generate GlcN6P²⁶. This reaction is the first step in the pathway for the production of UDP-GlcNAc, which is subsequently used in the process of cell wall biosynthesis. This biochemical pathway offers several possible metabolites whose concentrations might be sensed by a riboswitch for genetic control purposes.

Spontaneous cleavage of *glmS* RNA

In previous studies of riboswitches, we established the identity of the target compound of a given riboswitch by using an 'in-line probing' assay^{13–15,18,27,28}. This assay relies on the fact that unstructured regions of RNA typically undergo spontaneous cleavage more rapidly than the internucleotide linkages of RNA that are engaged in forming secondary and tertiary structures²⁹. In-line probing was performed with a 246-nucleotide RNA (termed 246-*glmS*) that carried most of the highly conserved portion of the intergenic region (IGR) preceding the *glmS* gene (Fig. 1b). Incubation of 246-*glmS* with candidate metabolites, including GlcN6P, revealed that a single cleavage fragment (Clv, Fig. 1c) dominated over the more modest amounts of products generated by typical spontaneous cleavage events.

This dominant cleavage event within the *glmS* RNA proceeds with a rate constant of $3 \times 10^{-3} \text{ min}^{-1}$ in the absence of GlcN6P, corresponding to a half-life for RNA cleavage of approximately 4 h (data not shown). Under similar reaction conditions, all known natural self-cleaving ribozymes⁴ exhibit half-lives of under 1 min. However, the *glmS* RNA cleaves more than 1,000-fold faster than that expected for spontaneous cleavage of an unconstrained RNA linkage³⁰ and more than tenfold faster than an RNA linkage that is constrained in a conformation that favours in-line nucleophilic attack^{29,31,32}. This atypically fast cleavage rate indicated that the RNA might constitute a novel self-cleaving ribozyme using multiple catalytic strategies to accelerate RNA cleavage. Furthermore, we speculated that conditions could be optimized such that the rate constant for *glmS* cleavage might become sufficient to be of biological significance.

A metabolite-dependent ribozyme

Examination of the small amounts of RNA that remained uncleaved at the dominant site provided some indications that the RNA also might undergo a conformational change when exposed to GlcN6P. This is evident by the apparent modulation of the intensities of several product bands corresponding to spontaneous cleavage events that occurred downstream of the ribozyme cleavage site (Fig. 1c). Therefore, we speculated that the *glmS* domain might function as a self-cleaving ribozyme whose activity depends on the binding of GlcN6P. There is precedence for modulation of

ribozymes in response to specific effector compounds. As part of their RNA-splicing activities, group I ribozymes catalyse cleavage of the first splice site junction at an accelerated rate when presented with guanosine or one of its phosphorylated derivatives, which serve as substrates for the reaction¹. In addition, we have demonstrated that numerous ligand-dependent ribozymes can be created by molecular engineering^{33–37}.

GlcN6P and several analogues (Fig. 2a) were used to determine whether the 246-*glmS* RNA functions as a metabolite-responsive ribozyme. The 246-*glmS* RNA undergoes rapid self-cleavage when incubated in the presence of 200 μM GlcN6P for 60 s, but remains significantly less active when incubated in the absence of metabolites or in the absence of Mg^{2+} (Fig. 2b). The level of cleavage product generated by the RNA when presented with GlcN6P compares favourably to that of other natural self-cleaving ribozymes, indicating that the 246-*glmS* RNA is indeed a ribozyme. Furthermore, most analogues of GlcN6P do not induce RNA cleavage to appreciable levels when present at 200 μM , indicating that the *glmS* ribozyme is selectively responsive to the metabolic product of the GlmS protein.

The specificity that the ribozyme exhibits for GlcN6P is demonstrated most markedly by the fact that the RNA discriminates against Glc6P, which differs from GlcN6P by a single functional group. One exception is the GlcN6P analogue wherein the phosphate is replaced by a sulphate (GlcN6S). This sulphate analogue of GlcN6P requires approximately 100-fold higher concentration to

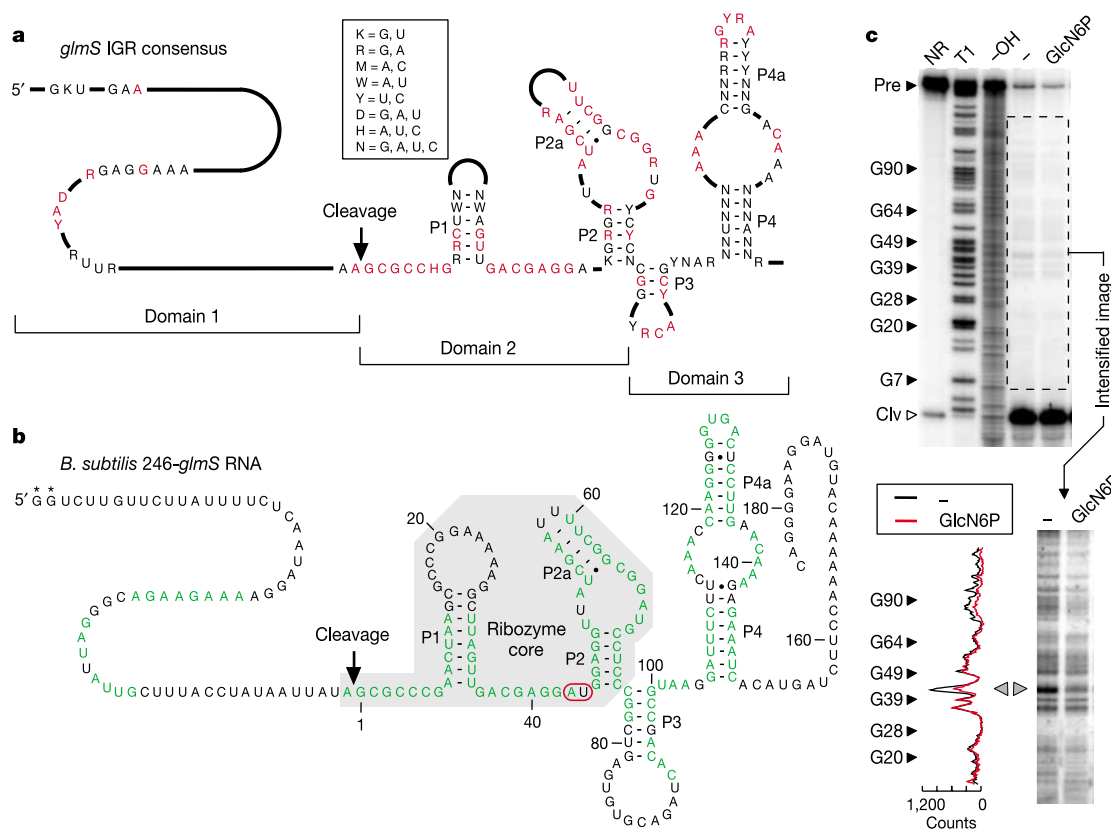


Figure 1 The *glmS* motif. **a**, Consensus sequence and secondary structure for the *glmS* motif based on 18 representatives (ref. 25 and data not shown). Nucleotides depicted in red and black reflect conservation levels of 95% and 80%, respectively. The arrow identifies the site of ribozyme-mediated cleavage. **b**, The *glmS* motif from *B. subtilis*. Green nucleotides match the consensus and asterisks denote nucleotides added to facilitate transcription *in vitro*. Circled nucleotides exhibit altered spontaneous cleavage rates with GlcN6P. The AUG start codon for *glmS* begins 245 nucleotides downstream of the ribozyme cleavage site. **c**, In-line probing analysis²⁹ of the 246-*glmS* RNA precursor

(Pre). NR, T1 and -OH identify no reaction, partial digestion with nuclease T1, and partial digestion with alkali, respectively. RNAs in the remaining two lanes were incubated in the absence (–) or presence of 1 mM GlcN6P. Although cleavage at the main site (CIV) is accelerated by 1,000-fold in the presence of GlcN6P (see Fig. 2), the 25-h incubation used in this assay allows the cleavage yield to be similar in the absence of GlcN6P. Enhanced view of the products of spontaneous cleavage events that occur downstream of the main cleavage site. Line plots from the two lanes are overlaid to highlight the change in cleavage yields at nucleotides 43 and 44.

trigger the ribozyme to the same extent (data not shown). Taken together, these findings suggest that the 246-*glmS* RNA carries a metabolite-dependent ribozyme that responds to GlcN6P with exceptional molecular specificity.

Characteristics of a ribozyme switch

The kinetic characteristics of the *glmS* ribozyme seem to be sufficient to serve as a riboswitch for the control of gene expression. A plot of the rate constant for ribozyme function versus the concentration of GlcN6P reveals that the rate constant for RNA cleavage becomes half maximal at a metabolite concentration of 200 μ M (Fig. 2c). Additional experiments are needed to determine whether this value reflects an apparent dissociation constant (K_d) wherein half the ribozymes are bound to the target metabolite, or whether ribozyme action follows more complex kinetic pathways. Although this metabolite concentration is higher than the apparent K_d values that have been determined for other riboswitch targets^{13–15,18,27,28}, the ribozyme becomes responsive to GlcN6P at concentrations as low as 200 nM.

The ribozyme also responds linearly to increasing GlcN6P concentrations, such that a tenfold increase in metabolite concentration produces a tenfold increase in ribozyme function. This linear dynamic range (DR) spans over 100-fold in ligand concentration and is consistent with the formation of a 1:1 complex between metabolite and ribozyme. Overall, the ribozyme exhibits a rate

acceleration of about 1,000-fold. Thus, saturation of the ribozyme with ligand reduces the half-life of the precursor RNA from approximately 4 h to less than 15 s.

The requirements for optimal ribozyme activity were established by determining the rate constants for 246-*glmS* under a variety of reaction conditions. In the presence of 10 mM Mg^{2+} , the ribozyme does not benefit from increasing ionic strength through the addition of spermidine or K^+ (data not shown). However, the ribozyme exhibits a near log-linear rate constant increase in response to increasing pH, with a plateau at a value of about 7 (Fig. 3a). These results suggest that the ribozyme can fold into its active structure, presumably involving the packing of negatively charged phosphates of the RNA chain in close proximity, when near physiological concentrations of Mg^{2+} are used.

The ribozyme does exhibit a high dependency on Mg^{2+} , even at a concentration of 200 mM K^+ , as evident by the marked reduction in the rate constant for self-cleavage at sub-physiological Mg^{2+} concentrations. Other natural ribozymes use metal ions for either folding or catalysis, and the substantial response to increasing Mg^{2+} with the *glmS* ribozyme suggests that multiple (at least four) divalent metal ions must be bound to achieve the maximum rate constant. A variety of divalent metal ions support ribozyme activity as well (Fig. 3b), which is characteristic of other natural self-cleaving ribozymes. In summary, the optimal activity for the *glmS* ribozyme occurs under near physiological conditions for Mg^{2+} and pH. However, further experiments are needed to determine the precise set of catalytic strategies^{31,32} that are used by this enzyme to accelerate RNA cleavage by more than 10-million-fold compared with the uncatalysed reaction.

Numerous ligand-responsive ribozymes that function as true allosteric enzymes have been created previously wherein ligand binding at a distal site favours the formation of a conformational state that permits catalysis to occur^{33–38}. Many engineered allosteric ribozymes can persist in an inactive state when the allosteric effector is absent, and then convert rapidly to the active state on addition of their target molecule. Similarly, we find that the *glmS* ribozyme

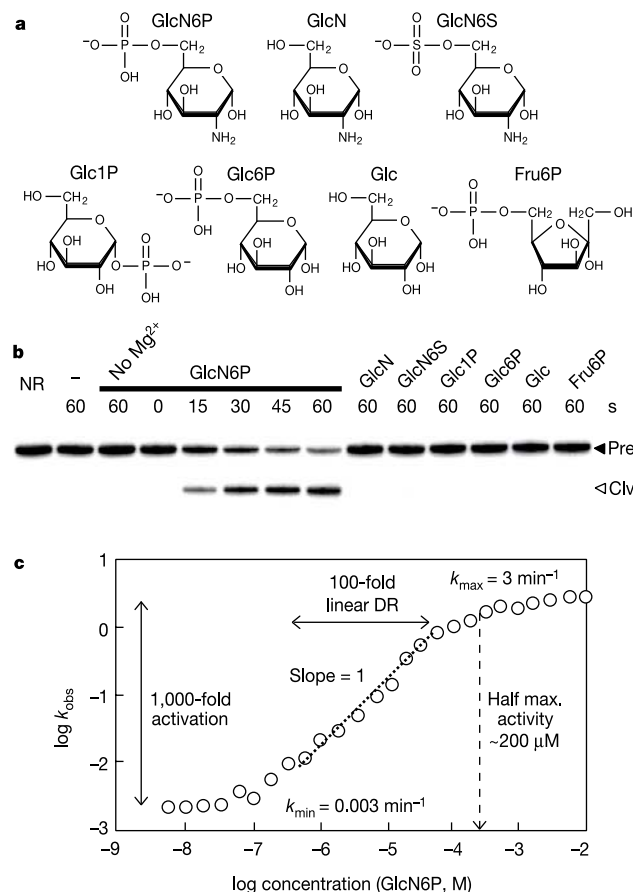


Figure 2 The *glmS* motif from *B. subtilis* serves as a metabolite-responsive ribozyme. **a**, Chemical structures and names of various compounds used to explore the molecular recognition characteristics of the *glmS* ribozyme. **b**, Ribozyme cleavage assays using the 246-*glmS* RNA and various compounds. 5' ³²P-labelled precursor (Pre) RNAs were incubated for the times indicated in the absence (–) or in the presence of 200 μ M effector as noted for each lane. **c**, Concentration-dependent activation of the 246-*glmS* ribozyme by GlcN6P. The maximum and minimum rate constants are denoted as k_{max} and k_{min} , respectively.

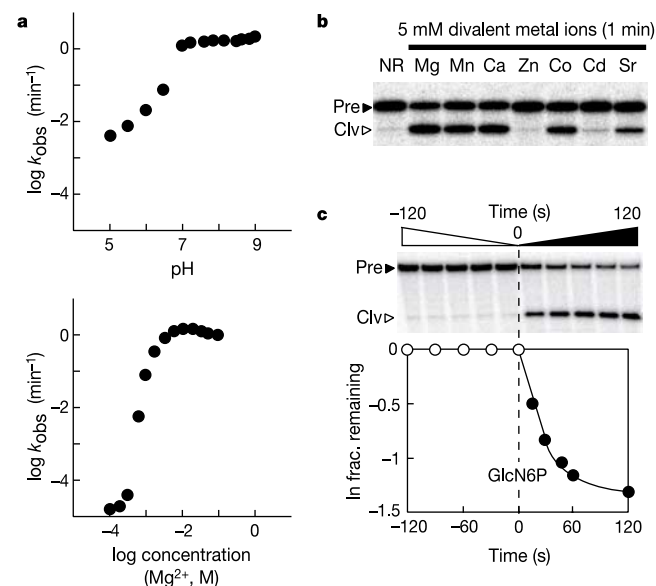


Figure 3 Optimization, divalent metal requirements and rapid metabolite induction of the *glmS* ribozyme. **a**, Plots depict the effects of pH and magnesium concentration on the rate constant for 246-*glmS* cleavage in 200 μ M GlcN6P under otherwise standard assay conditions. **b**, The cleavage activity of 246-*glmS* RNA in the presence of various divalent metal ions under otherwise standard conditions. NR indicates no reaction whereas Pre and Clv identify precursor and 5' cleavage products, respectively. **c**, Rapid activation of the *glmS* ribozyme was carried out under standard conditions, wherein GlcN6P was added to 100 μ M at time 0.

responds instantaneously to the addition of GlcN6P (Fig. 3c), indicating that the RNA serves as a rapid switch that can form its active structure without the need for a forceful denaturation and annealing step. This finding suggests that the *glmS* mRNA might persist for extended periods of time in the cell without undergoing significant ribozyme cleavage, and then respond rapidly to rising concentrations of GlcN6P.

We are not yet able to specify whether this RNA functions as a true allosteric ribozyme, or whether GlcN6P might take on a chemical role in RNA cleavage by serving directly as a part of the active site. Engineered allosteric ribozymes typically can be dissected such that the separate ribozyme and ligand-binding domains can function independently³⁷. At this time, it is not clear how this dissection might be achieved with the *glmS* ribozyme. We have determined that Glc6P does not function as a competitive inhibitor of the ribozyme (data not shown), indicating that this analogue of GlcN6P is inactive because it fails to occupy the metabolite binding site and not (only) because it might lack a functional group that is critical for catalysis.

A minimal ribozyme switch

Determination of the minimum active ribozyme domain was achieved by testing a series of truncated RNA constructs (Fig. 4a). The activities of five RNA constructs with successive single-nucleotide deletions flanking the ribozyme cleavage site were determined in the presence and absence of 200 μ M GlcN6P (Fig. 4a). To

facilitate preparation by transcription *in vitro*, each RNA carried a G residue at its 5' terminus in addition to the natural nucleotide sequence as indicated for each construct. RNAs that retained at least one nucleotide upstream of the cleavage site exhibited ribozyme activity, albeit at a reduced rate.

The ribozyme was also minimized by deletion of nucleotides at the 3' terminus. Initial experiments to define the 3' boundary made use of a nested set of RNAs that were generated by partial alkaline hydrolysis. RNAs that were truncated downstream of the P4 stem appear to lose some activity, whereas RNAs that were truncated beyond the junction between P3 and P4 lose even more activity (data not shown). Metabolite-responsive ribozyme activity was not completely lost until truncations were made upstream of nucleotide 75 at the base of P2 (Fig. 4a).

These findings indicate that only domain 2 of the *glmS* element (Fig. 1a) is required for catalytic function *in vitro* despite the fact that considerable portions of the conserved sequences reside in domains 1 and 3. Although the role of domain 1 remains unresolved, the P3 and P4 elements in part support greater activity of the *glmS* ribozyme. The functional core of the ribozyme does not appear to form conventional base-pair interactions that constrain the structure and location of target linkages as in most other ribozymes. Possibly, the highly conserved nucleotides that occur immediately downstream of the cleavage site are involved in tertiary structure that constrains the labile linkage for catalysis.

Cleavage mechanism of the *glmS* ribozyme

Of the eight classes of ribozymes identified previously, four catalyse RNA cleavage by promoting internal phosphoester transfer (Fig. 4b). This mechanism proceeds when the 2' oxygen atom at the cleavage site (I) performs a nucleophilic attack on the adjacent phosphorus centre (II). The newly generated termini of the cleavage products carry 2',3'-cyclic phosphate (III) and 5' hydroxyl (IV) moieties. The four self-cleaving ribozymes that catalyse this reaction are the smallest of the natural ribozymes⁴, which suggests that it is relatively straightforward to form an active site that promotes this chemical transformation. Consistent with this hypothesis is the fact that spontaneous RNA cleavage by internal phosphoester transfer occurs with a speed that is nearly 1-million-fold faster than that for external phosphoester transfer, including direct hydrolysis of a phosphodiester linkage^{29,30}. Given the relative ease of promoting RNA cleavage by internal phosphoester transfer, and given the relatively small size of the *glmS* ribozyme core, we speculated that the *glmS* ribozyme might also use this mechanism for RNA cleavage.

If the *glmS* ribozyme reaction proceeds via internal phosphoester transfer (Fig. 4b), then the 5' cleavage fragment should carry a 2',3'-cyclic phosphate (III). Under typical assay conditions used herein, the cyclic phosphate products are expected to degrade slowly to form a mixture of the 2' and 3' phosphate products in a hydrolytic process that can be accelerated by treatment with acid³⁸. Because the cyclic phosphate form can be resolved from the open form by gel electrophoresis, we compared the gel mobilities of the acid-treated and untreated 5' cleavage fragments from a truncated *glmS* ribozyme that carried only five nucleotides upstream of the cleavage site (Fig. 4c). Without acid treatment, the cleavage fragment has a mobility that precisely matches that of the fragment generated by partial alkaline digestion, which is known to generate 2',3'-cyclic phosphate products. On acid treatment, this putative cyclic phosphate product is destroyed and a second product band emerges that corresponds in mobility to that expected for the open phosphate forms. These findings provide preliminary support for the hypothesis that the *glmS* ribozyme catalyses RNA cleavage by means of internal phosphoester transfer.

The *glmS* ribozyme is a genetic switch

The robust cleavage activity of the full-length *glmS* motif and its selective response to GlcN6P are strong indications that the

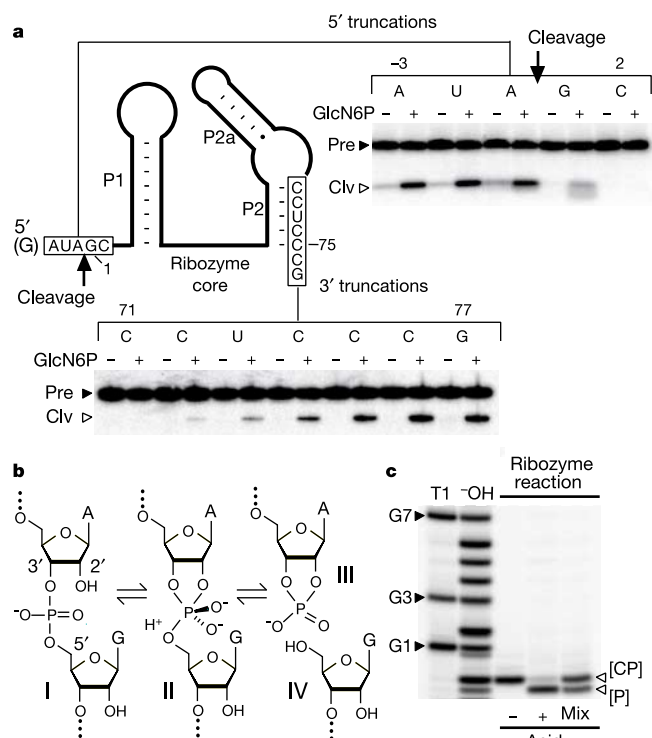


Figure 4 Ribozyme boundaries and cleavage site chemistry. **a**, Activities of various *glmS* ribozyme deletion mutants in the absence (–) or presence (+) of 200 μ M GlcN6P. The 5' end deletion constructs are initiated with a non-wild-type guanosine (G) residue to facilitate RNA preparation by transcription *in vitro*. **b**, Phosphoester transfer mechanism for RNA cleavage used by natural self-cleaving ribozymes. **c**, High-resolution mapping of the *glmS* 5' cleavage product. The dominant 5' cleavage product in the absence (–) of acid treatment carries a 2',3'-cyclic phosphate group and is denoted as [CP]. Acid treatment (+) after ribozyme action produces a distinct product band corresponding to 5' cleavage fragments with either a 2'– or a 3'–phosphate group denoted as [P], and the mobility difference between the two products is confirmed when combined in a single lane (mix). Other notations are as described in the legend to Fig. 1c.

metabolite-responsive ribozyme is of biological significance. Furthermore, the ribozyme consensus sequence remains highly conserved among Gram-positive organisms, and is always located immediately upstream of the *glmS* coding region. The most likely purpose for the metabolite-dependent action of the ribozyme (at least in part) is to reduce *glmS* gene expression in response to increasing GlcN6P concentrations. This hypothesis was explored by generating a series of mutants and testing their functions as self-cleaving ribozymes *in vitro* and as genetic control elements *in vivo*. Some mutations were introduced into portions of the RNA that were expected to have a minimal impact on ribozyme function, whereas other mutations were placed in positions that were expected to be more disruptive (Fig. 5a).

Modest reductions in rate constant for ribozyme cleavage were observed with the mutants M1–M7 (Fig. 5a). These findings are in accordance with the modest loss in the extent of ribozyme cleavage on deletion of the conserved nucleotides upstream of the cleavage site, downstream of nucleotide 75, and within the non-conserved loop of P1 (Fig. 4a). In contrast, the self-cleavage activities of mutants M8 and M9 are not detected, indicating that the identity of the nucleotide located immediately 3' to the labile linkage is essential for catalytic activity. Notably, we observe an inverse relationship between ribozyme activity *in vitro* and the *in vivo* activity of a β -galactosidase reporter gene that is fused to the various *glmS* IGR mutants (Fig. 5b). Specifically, mutants that exhibit near wild-type ribozyme activity also exhibit low levels of β -galactosidase activity, which are similar to those measured when the wild-type *glmS* IGR is fused to the reporter. In contrast, the M8 and M9

reporter constructs exhibit more than 20-fold greater β -galactosidase activity, suggesting that ribozyme activity represses the *glmS* gene.

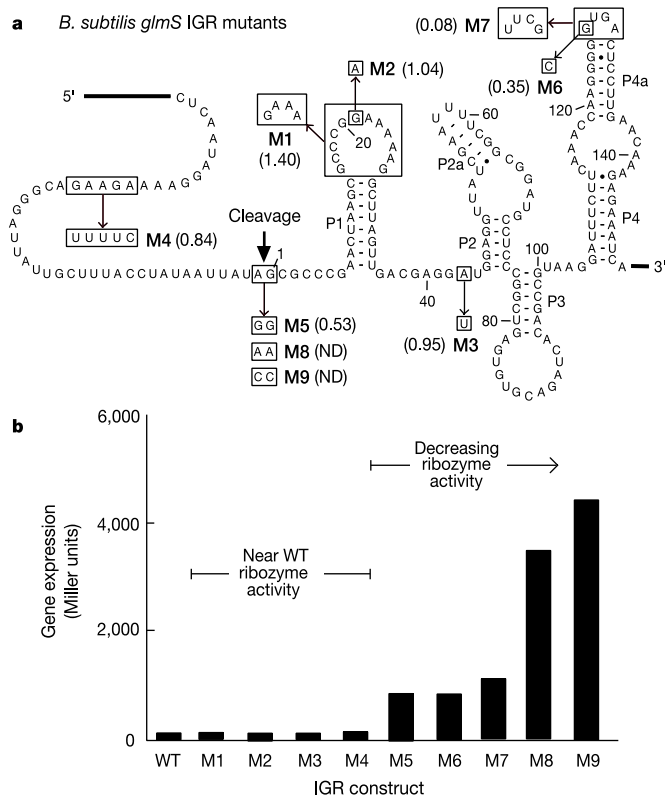
Discussion

It has been more than a decade since the last new ribozyme motif was reported³⁹, which suggests that there are very few different types of catalytic RNAs present in extant organisms. However, it is likely that life passed through a period in evolution wherein RNA had to serve a much larger role as both catalyst and receptor for the production and sensing of metabolic products⁴⁰. The examples of natural riboswitches¹⁷ and engineered allosteric ribozymes^{8–10} provide evidence that RNA has sufficient functional potential to maintain a complex metabolic state without the need for proteins. Moreover, the demonstration that the *glmS* RNA operates as a metabolite-responsive ribozyme that controls gene expression proves that the functional capabilities possibly exploited by ancient life forms are also adequate to fulfil important roles in genetic regulation even in modern organisms.

The riboswitches discovered previously do not use a ribozyme as part of their mechanism of genetic control¹⁷. We speculate that, in most instances, a riboswitch formed by combined metabolite binding and self-cleaving ribozyme domains would require a more complex RNA sequence and structural motif compared to a related, but non-catalytic, riboswitch. Presumably the formation of both an aptamer for metabolite binding and an active site for catalysis would need to be more complex and remain relatively constant through evolution compared with a riboswitch that binds a target compound and controls gene expression via simple structural rearrangement. If true, then we would expect that an allosteric ribozyme whose purpose is exclusively to control an adjacent gene would lose its catalytic activity through evolution, but retain its genetic control function in the form of a simpler riboswitch. We have determined that the related *glmS* RNAs from *Bacillus cereus* and *Thermoanaerobacter tengcongensis* also undergo GlcN6P-dependent self-cleavage with rate constants that are comparable to that of the *B. subtilis* ribozyme (data not shown). Thus, the widespread use of the *glmS* ribozyme in Gram-positive organisms implies that its catalytic function might be necessary for biomolecular processes beyond just controlling the *glmS* gene.

It is possible that the *glmS* ribozyme has not atrophied through evolution because its ability to bind GlcN6P might be inextricably coupled to its catalytic activity, such that mutations disrupting catalysis also disrupt metabolite binding. Given the considerations described above, it is interesting to note that some sequence conservation is present in domain 1 of the *glmS* element (Fig. 1a), despite the fact that the sequence can be deleted (Fig. 4a) or mutated (Fig. 5a) with little effect on ribozyme function. One possibility is that the ribozyme is required to control the *glmS* gene as well as release domain 1 as a 5' cleavage fragment. Once liberated from the *glmS* mRNA, this non-coding RNA might then serve as a signal of adequate GlcN6P concentrations by influencing other cellular processes.

The *glmS* ribozyme responds to GlcN6P and exhibits a high level of discrimination against even closely related metabolites. This characteristic is shared with other riboswitches^{13–15,18,21,23,27}, which similarly must provide the cell with a means to control gene expression with great precision in a complex mixture of biological molecules. GlcN6P is a compound that is crucial to sugar metabolism and cell wall biosynthesis in Gram-positive organisms, and it is an essential metabolite for organisms in all three phylogenetic kingdoms. Thus, GlcN6P joins an expanding list of fundamental metabolites that are recognized by RNA for genetic control purposes^{17,18}. Given that there are a number of new RNA motifs in prokaryotes whose functions have yet to be assigned²⁵, this list may expand substantially to include other metabolites that are central to major metabolic pathways.



Methods

Chemicals and oligonucleotides

GlcN6P and its analogues glucosamine (GlcN), glucosamine-6-sulphate (GlcN6S), glucose-1-phosphate (Glc1P), glucose-6-phosphate (Glc6P), glucose (Glc) and fructose-6-phosphate (Fru6P) were purchased from Sigma. Synthetic DNAs were purchased and purified as described previously¹¹.

RNA analyses

In-line probing was conducted with 5' ³²P-labelled RNAs using methods similar to those described previously^{13–15}. Rate constants for ribozyme cleavage were determined using trace amounts of 5' ³²P-labelled RNA incubated under in-line probing conditions, standard conditions (50 mM Tris-HCl (pH 7.5 at 23 °C), 10 mM MgCl₂ and 200 mM KCl), or as otherwise described for each experiment. Reactions also were conducted in the absence or presence of GlcN6P as noted, and were incubated at 23 °C. Aliquots were withdrawn at various times and terminated by the addition of either four volumes of formamide loading buffer (95% formamide, 10 mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue), or with a urea-containing loading buffer that was supplemented with 100 mM EDTA. The products were analysed using denaturing polyacrylamide gel electrophoresis (PAGE) and the rate constants for RNA cleavage were determined as described previously³². Data points collected at later stages of the reaction (beyond 75% cleavage) begin to exhibit nonlinearity and are not included when establishing the rate constant for the initial stage of the reaction. Under standard conditions in the presence of 50 μM GlcN6P, 10% of 246-*glmS* RNA remains uncleaved after a 90-min incubation at 23 °C.

Ribozyme truncations

Initial experiments to establish the approximate 3' boundary for the ribozyme domain were carried out by simultaneously examining the activities of a nested set of RNAs. Specifically, 5' ³²P-labelled precursor RNAs subjected to partial alkaline digestion were incubated in the presence of 200 μM GlcN6P under standard assay conditions. RNAs within this population that fail to retain near full ribozyme activity persist as uncleaved RNAs, and thus appear as bands upon separation of reaction products by denaturing 6% PAGE. The 5' deletions were made with a parent construct that extends 183 nucleotides downstream of the cleavage site. The 3' deletions were made with a parent construct that extends 63 nucleotides upstream of the cleavage site (Fig. 1b). Specific truncated RNAs were generated by *in vitro* transcription using T7 RNA polymerase and the corresponding template DNAs. Each DNA template was prepared by polymerase chain reaction (PCR) amplification of the IGR sequence with the appropriate primers. Transcribed RNAs were purified by denaturing PAGE, dephosphorylated and 5' ³²P-labelled as described previously¹¹. Ribozyme assays were conducted under standard assay conditions.

Ribozyme cleavage site analysis

A 5' ³²P-labelled *glmS* RNA construct containing four nucleotides upstream of the cleavage site (5' ³²P-GAUA plus 100 nucleotides downstream from the cleavage site) was incubated for 6 min at 23 °C under standard conditions in the presence of 200 μM GlcN6P. Acid treatment of the reaction products consisted of a 2 h incubation at 23 °C in the presence of 25 mM HCl³⁸. The sample was neutralized by addition of NaOH to a final concentration of 25 mM before analysis by denaturing 20% PAGE.

Construction of ribozyme mutants and reporter fusions

Mutations were generated from plasmid pDG1661 carrying the entire *glmS* IGR fused upstream from a *lacZ* reporter gene, and examined using methods similar to those reported previously²⁷. Briefly, a construct encompassing nucleotides –379 to +39 relative to the translation start site of *glmS* was prepared using primers that generated *EcoRI* and *BamHI* restriction sites upon PCR amplification of *B. subtilis* chromosomal DNA (strain 1A40). The product was cloned into pDG1661 using these restriction sites to place the *glmS* IGR immediately upstream of the *lacZ* reporter gene. Templates for preparation of mutant RNAs for *in vitro* studies were created by PCR from mutated plasmids using the appropriate primers. All DNA constructs were integrated into the *B. subtilis* chromosome (strain 1A40) for *in vivo* studies.

In vivo expression assays

Cells were grown with shaking at 37 °C in defined media (0.5% (w/v) glucose, 2 g l^{–1} (NH₄)₂SO₄, 18.3 g l^{–1} K₂HPO₄·H₂O, 6 g l^{–1} KH₂PO₄, 1 g l^{–1} sodium citrate, 0.2 g l^{–1} MgSO₄·H₂O, 5 μM MnCl₂, 0.5 mM CaCl₂, 50 μg ml^{–1} tryptophan, 50 μg ml^{–1} methionine, 50 μg ml^{–1} lysine with chloramphenicol (5 μg ml^{–1})). Growth under glucose-limiting conditions was established by incubation under routine growth conditions in defined media to an A₅₉₅ of 0.1 at which time the cells were pelleted by centrifugation and resuspended in defined media without glucose. Cultures were incubated for an additional 4 h before β-galactosidase assays were carried out using a protocol similar to that described previously⁴¹.

Received 5 December 2003; accepted 19 January 2004; doi:10.1038/nature02362.

- Cech, T. R. Self-splicing of group I introns. *Annu. Rev. Biochem.* **59**, 543–568 (1990).
- Frank, D. N. & Pace, N. R. Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.* **67**, 153–180 (1998).
- Michel, F. & Feral, J. Structure and activities of group II introns. *Annu. Rev. Biochem.* **64**, 435–461 (1995).
- Doherty, E. A. & Doudna, J. A. Ribozyme structures and mechanisms. *Annu. Rev. Biochem.* **69**, 597–615 (2000).

- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
- Breaker, R. R. *In vitro* selection of catalytic polynucleotides. *Chem. Rev.* **97**, 371–390 (1997).
- Osborne, S. E. & Ellington, A. D. Nucleic acid selection and the challenge of combinatorial chemistry. *Chem. Rev.* **97**, 349–370 (1997).
- Soukup, G. A. & Breaker, R. R. Allosteric nucleic acid catalysts. *Curr. Opin. Struct. Biol.* **10**, 318–325 (2000).
- Breaker, R. R. Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.* **13**, 31–39 (2002).
- Silverman, S. K. Rube Goldberg goes (ribo)nuclear? Molecular switches and sensors made from RNA. *RNA* **9**, 377–383 (2003).
- Seetharaman, S., Zivarts, M., Sudarsan, N. & Breaker, R. R. Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nature Biotechnol.* **19**, 336–341 (2001).
- Ptashe, M. & Gann, A. *Genes & Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002).
- Nahvi, A. *et al.* Genetic control by a metabolite-binding mRNA. *Chem. Biol.* **9**, 1043–1049 (2002).
- Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).
- Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. *Proc. Natl Acad. Sci. USA* **99**, 15908–15913 (2002).
- Mironov, A. S. *et al.* Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**, 747–756 (2002).
- Winkler, W. C. & Breaker, R. R. Genetic control by metabolite-binding riboswitches. *ChemBiochem* **4**, 1024–1032 (2003).
- Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**, 577–586 (2003).
- Gusarov, I. & Nudler, E. The mechanism of intrinsic transcription termination. *Mol. Cell* **3**, 495–504 (1999).
- Yarnell, W. S. & Roberts, J. W. Mechanism of intrinsic transcription termination and antitermination. *Science* **284**, 611–615 (1999).
- Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Struct. Mol. Biol.* **11**, 29–35 (2004).
- Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. & Gelfand, M. S. Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.* **20**, 44–50 (2004).
- Sudarsan, N., Barrick, J. E. & Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* **9**, 644–647 (2003).
- Kubodera, T. *et al.* Thiamine-regulated gene expression of *Aspergillus oryzae* thiA requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR. *FEBS Lett.* **555**, 516–520 (2003).
- Barrick, J. E. *et al.* New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl Acad. Sci. USA* (submitted).
- Milewski, S. Glucosamine-6-phosphate synthase: the multi-facets enzyme. *Biochim. Biophys. Acta* **1597**, 173–192 (2002).
- Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Struct. Biol.* **10**, 701–707 (2003).
- Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S. & Breaker, R. R. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* **17**, 2688–2697 (2003).
- Soukup, G. A. & Breaker, R. R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**, 1308–1325 (1999).
- Li, Y. & Breaker, R. R. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. *J. Am. Chem. Soc.* **121**, 5364–5372 (1999).
- Emilsson, G. M., Nakamura, S., Roth, A. & Breaker, R. R. Ribozyme speed limits. *RNA* **9**, 907–918 (2003).
- Breaker, R. R. *et al.* A common speed limit for RNA-cleaving ribozymes and deoxyribozymes. *RNA* **9**, 949–957 (2003).
- Tang, J. & Breaker, R. R. Rational design of allosteric ribozymes. *Chem. Biol.* **4**, 453–459 (1997).
- Soukup, G. A. & Breaker, R. R. Engineering precision RNA molecular switches. *Proc. Natl Acad. Sci. USA* **96**, 3584–3589 (1999).
- Koizumi, M., Soukup, G. A., Kerr, J. Q. & Breaker, R. R. Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nature Struct. Biol.* **6**, 1062–1071 (1999).
- Breaker, R. R. Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.* **13**, 31–39 (2002).
- Soukup, G. A., DeRose, E. C., Koizumi, M. & Breaker, R. R. Generating new ligand-binding RNAs by affinity maturation and disintegration of allosteric ribozymes. *RNA* **7**, 524–536 (2001).
- Abrash, H. I., Cheung, C.-C. S. & Davis, J. C. The nonenzymic hydrolysis of nucleoside 2', 3'-phosphates. *Biochemistry* **6**, 1303 (1967).
- Saville, B. J. & Collins, R. A. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora mitochondria*. *Cell* **61**, 685–696 (1990).
- Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).
- Miller, J. H. *A Short Course in Bacterial Genetics* 72 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1992).

Acknowledgements We thank K. Corbino for generating the *glmS* IGR sequence alignments, and members of the Breaker laboratory for discussions. This work was supported by grants from the NIH and the NSF. R.R.B. is also grateful for support from the Yale Liver Center and the David and Lucile Packard Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.R.B. (ronald.breaker@yale.edu).

Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution

Zhenfeng Liu¹, Hanchi Yan¹, Kebin Wang², Tingyun Kuang², Jiping Zhang¹, Lulu Gui¹, Xiaomin An¹ & Wenrui Chang¹

¹National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, People's Republic of China

²Laboratory of Photosynthesis and Environmental Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Xiangshan, Beijing 100093, People's Republic of China

The major light-harvesting complex of photosystem II (LHC-II) serves as the principal solar energy collector in the photosynthesis of green plants and presumably also functions in photoprotection under high-light conditions. Here we report the first X-ray structure of LHC-II in icosahedral proteoliposome assembly at atomic detail. One asymmetric unit of a large R32 unit cell contains ten LHC-II monomers. The 14 chlorophylls (Chl) in each monomer can be unambiguously distinguished as eight Chla and six Chlb molecules. Assignment of the orientation of the transition dipole moment of each chlorophyll has been achieved. All Chlb are located around the interface between adjacent monomers, and together with Chla they are the basis for efficient light harvesting. Four carotenoid-binding sites per monomer have been observed. The xanthophyll-cycle carotenoid at the monomer–monomer interface may be involved in the non-radiative dissipation of excessive energy, one of the photoprotective strategies that have evolved in plants.

Light harvesting is the primary process in photosynthesis. In green plants, the function of harvesting solar energy is fulfilled by a series of light-harvesting complexes in the thylakoid membrane of chloroplasts. LHC-II, the most abundant integral membrane protein in chloroplasts, exists as a trimer and binds half of the thylakoid chlorophyll molecules. Every monomeric LHC-II comprises a polypeptide of about 232 amino-acid residues, 13–15 Chla and Chlb molecules¹, 3–4 carotenoids² and one tightly bound phospholipid³. Besides the light-harvesting function, LHC-II has also been shown to function in the non-radiative dissipation of excess excitation energy formed under high-light conditions^{4,5}. It has a crucial role in minimizing the damaging effects of excess light by operating this photoprotective mechanism as light intensity becomes increasingly saturating. Moreover, LHC-II also takes part in regulating the distribution of excitation energy to photosystems II and I (ref. 6).

The structure of LHC-II from pea has been determined by electron crystallography at 3.4 Å resolution parallel to the membrane plane, and at about 4.9 Å resolution perpendicular to this plane⁷. This model revealed some basic structural features of LHC-II, including three transmembrane α -helices (helices A, B and C) and a short amphipathic helix (helix D), 12 chlorophyll tetrapyrroles with roughly determined locations and orientations, and two carotenoids. A more detailed structural picture of LHC-II, with an unambiguous determination of the identity of the chlorophylls (Chla or Chlb) and the orientation of their transition dipole moments, would be beneficial for a better understanding of the basic functional mechanism of LHC-II. We have obtained this information by solving the structure of LHC-II at higher resolution using X-ray crystallography.

In our model of the X-ray structure of LHC-II at 2.72 Å resolution, we provide the basis for investigating quantitatively the underlying mechanism of the light-harvesting process and its adjustment in LHC-II. We also reveal for the first time an elegant arrangement of membrane proteins in the icosahedral proteoliposome assembly, and show that membrane proteins can be crystallized in a way that differs from those described in ref. 8.

Structure determination

Crystallization of trimeric LHC-II isolated from *Spinacia oleracea* is described briefly in the Methods. Structure determination of

LHC-II organized in an icosahedral particle includes initial phasing by the single isomorphous replacement (SIR) method plus phase refining and extending by the real-space averaging method⁹. Data collection, phasing and refinement statistics are listed in Table 1.

The high-quality electron density map enabled us to trace 94% of

Table 1 Data collection, phasing and refinement statistics

Data set	Native (1)	Native (2)	K ₂ HgI ₄ (derivative)
Resolution (Å)	50–3.5	25–2.7	50–3.5
R _{merge} *	0.087 (0.112)	0.082 (0.368)	0.100 (0.131)
Completeness (%)	97.9 (95.5)	90.8 (79.6)	80.5 (83.2)
<I/σ>	6.1 (2.1)	14.4 (2.5)	6.4 (3.3)
SIR phasing statistics			
No. of heavy atom sites	10		
R _{cullis} (centric/accentric)†	0.66/0.69		
Phasing power (centric/accentric)†	1.13/1.67		
Resolution (Å)	15–5.0		
FOM‡	0.36		
Phase refinement and extension statistics			
Resolution (Å)	10–2.72		
FOM‡	0.922 (0.801)		
Correlation coefficient‡	0.967 (0.765)		
R factor‡	0.128 (0.468)		
Structure refinement statistics			
Resolution	10–2.72		
Reflections (working set)	179,170		
Reflections (test set)	9,326		
R _{work} /R _{free} (%)	19.4/22.1		
r.m.s.d. bond length (Å)	0.0124		
r.m.s.d. bond angle (°)	2.257		
Coordinate error (Å)§			
Luzzati§	0.28		
SigmaA§	0.28		
Number of non-hydrogen atoms			
Protein	16,619		
Cofactors	11,720		
Water	699		

Numbers in parentheses correspond to values in the highest resolution shell.

* $R_{\text{merge}} = \sum_i \sum_h |I_{ih} - \langle I_h \rangle| / \sum_i \sum_h I_{ih}$, where h are unique reflection indices, I_{ih} are intensities of symmetry-related reflections and $\langle I_h \rangle$ is the mean intensity. Reflections with $I > 2.0 \times \sigma_I$ in native (1) and derivative data sets were used in the R_{merge} calculation, whereas the $\sim 3.0 \times \sigma_I$ cutoff was applied in the native (2) data set.

†FOM (figure of merit), R_{cullis} (centric) and phasing power determined by programs from the CCP4 suite (see Methods).

‡Correlation coefficient = $\sum_h (\langle F_o \rangle - |F_c|) / (\langle F_o \rangle - |F_c|) / [\sum_h (\langle F_o \rangle - |F_c|)^2 \sum_h (\langle F_o \rangle - |F_c|)^2]^{1/2}$, R factor = $\sum_h |F_o - F_c| / \sum_h F_o$, where h are the unique reflection indices, F_o are the observed structure factors and F_c are the structure factors calculated from inversion of the non-crystallographic symmetry-averaged map.

§From Luzzati plot and SigmaA analysis, as determined with CNS (see Methods).

the 232 amino acids and accurately locate the 14 chlorophylls and 4 carotenoids within one monomeric LHC-II. For the 14 chlorophylls, assignment of the orientation of the Q_x and Q_y transition dipolar moments was accomplished by proper positioning of the chlorophyll head groups. Ten chlorophylls were modelled with complete phytyl chains, but phytyl chains for the remaining four chlorophylls could be only partially modelled. With the help of $2F_o - F_c$ and $F_o - F_c$ electron-density maps (Fig. 1a, b), all 14 chlorophylls were unambiguously characterized as eight Chla and six Chlb. The resulting Chla/b molar ratio of 1.33 is consistent with the value determined by earlier biochemical analyses^{1,2}. Three carotenoids were identified as two luteins and one neoxanthin, and the fourth member was interpreted as a mixed density involving the xanthophyll-cycle carotenoids. In addition, two lipids, one detergent molecule and about 70 water molecules per monomer have been positioned. Figure 1c, d shows two regions of the electron-density map calculated with the phases at 2.72 Å resolution.

Icosahedral proteoliposome and crystal packing

In the $T = 1$ icosahedral particle (Fig. 2a), 20 LHC-II trimers are organized in a closed '532' point group symmetry, with their central C_3 axis serving as the icosahedral C_3 axis and oriented radially towards the sphere centre. One C_3 axis and two C_2 axes of the icosahedron superpose with the crystallographic axes. These trimers form a spherical shell with an outer diameter of about 261 Å and an inner diameter of about 160 Å. They are oriented in the shell with their flat luminal surface facing the interior of the sphere and the less flat stromal surface facing outwards, taking part in the contacts with other particles in the crystal. The interactions between two adjacent trimers are mediated mainly by two digalactosyl diacylglycerol (DGDG) molecules and two pairs of chlorophylls through van der Waals contacts. They are all located near the icosahedral C_2 axis. The digalactosyl head group of each DGDG is simultaneously hydrogen bonded to the luminal-surface amino acids from two

adjacent trimers, functioning as a bridge. The hydrophobic fatty-acid chains of DGDG extend into the membrane interior, interacting with hydrophobic residues and pigments of LHC-II. Other lipids that are expected to fill the gaps between LHC-II trimers and to form a spherical lipid-bilayer vesicle are mostly disordered.

In the crystal lattice (Fig. 2b), LHC-IIs are assembled and packed in a manner different from those in 'Type I' and 'Type II' three-dimensional (3D) crystals of membrane proteins as originally proposed in ref. 8. The LHC-DGDG proteoliposomes assume the shape of closed spheres, presumably originating from curved, small patches of two-dimensional (2D) membrane-protein crystals. Both the outer and inner surfaces of each proteoliposome are hydrophilic. The contacts between two proteoliposomes in the crystal lattice are polar interactions provided by the hydrophilic stromal surfaces of LHC-II. The hydrophobic intramembranous surfaces of LHC-II trimers are sheltered from crystal packing by the hydrophobic chains of lipids. We categorize this novel kind of 3D crystal as a 'Type III' membrane-protein crystal.

The apoprotein and LHC-II trimer

The polypeptide main chain of each monomeric LHC-II was continuously traced from Ser 14 to Gly 231. The secondary structure model of spinach LHC-II reported here (Fig. 3a) is similar to the electron crystallographic model of LHC-II from pea⁷. Between the primary structures of spinach and pea LHC-II, 89% of the 232 amino acids are conserved. However, deviations in the residue range, length, turns and orientation between helices in the two species were observed⁷ (Fig. 3a). We also found a typical amphipathic short 3_{10} -helix located in the BC loop region and named it helix E. Helix E has a length close to that of helix D and is related to helix D by the internal pseudo- C_2 axis. It is inclined with respect to the membrane plane by an angle of about 30°. In the following EC loop, the polypeptide folds into two short antiparallel strands that are stabilized by an inter-strand ionic pair (Asp 111–His 120) and some hydrogen bonds.

The basic structural and functional unit of LHC-II is the trimer. The whole trimerization region covers the amino-terminal domain, the carboxy terminus, the stromal end of helix B, several hydrophobic residues from helix C and also the pigments and lipid bound to these parts of the polypeptide chain (Fig. 3b). Chla 614, Chla 613, xanthophyll-cycle carotenoid, phosphatidylglycerol (PG), Chlb 601 and Chla 602 from one monomer together with Chlb 607, Chlb 609 and Chla 603 from the neighbouring monomer line up from the periphery of the trimer to the core region near the central C_3 axis at

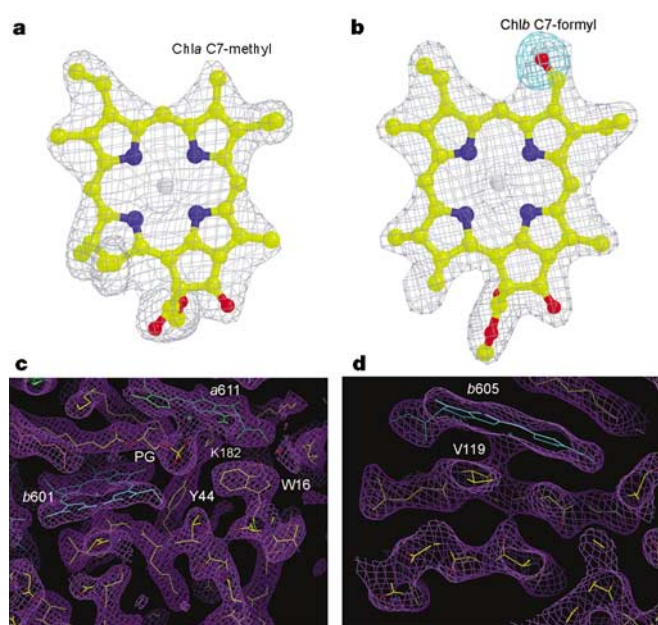


Figure 1 Electron-density map at 2.72 Å resolution. **a**, Chla and **b**, Chlb. Grey cage, $2F_o - F_c$ density ($1.5 \times \sigma$ level); cyan cage, $F_o - F_c$ density ($4.0 \times \sigma$ level). No residual $2F_o - F_c$ or $F_o - F_c$ density appears beside Chla C7-methyl, while strong $2F_o - F_c$ and $F_o - F_c$ densities show up at the position of Chlb C7-formyl if it is omitted. **c**, N-terminal region including binding sites for a Chlb (cyan) and a phospholipid coordinated to a Chla (green). **d**, Two antiparallel polypeptide strands in the EC loop region with one Chlb bound. In **c** and **d**, $2F_o - F_c$ densities ($1.5 \times \sigma$ level) are shown as a purple cage.

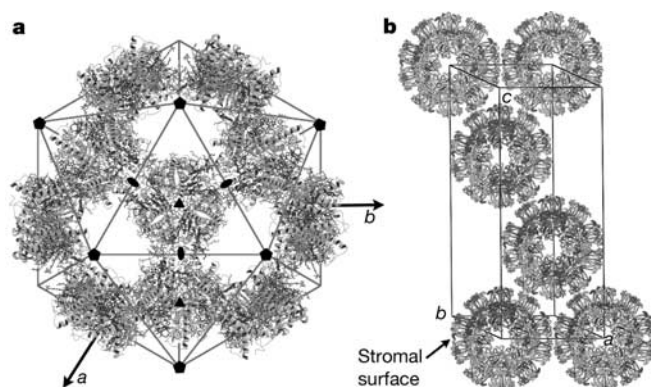


Figure 2 Organization and packing of the icosahedral particles. **a**, Schematic drawing of one-half of the LHC-II-DGDG proteoliposome viewed along the c axis of the hexagonal cell. **b**, Packing diagram of 'Type III' membrane-protein crystal, showing the contacts between icosahedral spherical particles in the hexagonal cell. Prosthetic groups are omitted for clarity. The N-terminal domain and AC loop region located at the stromal surface are involved in the crystal packing.

the interface between monomers, forming extensive hydrophobic interactions. Six Chla (Chla 602 and Chla 603 from each monomer) constitute the core of the trimer. Our observations directly reveal the structural role of PG in stabilizing the LHC-II trimer and clearly indicate that hydrophobic interactions dominate the associations between monomers within a trimer. It was shown that removal of the first 49 or 51 amino-acid residues of the polypeptide by proteolytic cleavage led to loss of PG and complete dissociation of the trimer into monomers, and that hydrolysis of the PG by phospholipase A₂ has a similar effect in breaking down the LHC-II trimer³.

Chlorophyll-binding sites

In a crystallographic asymmetric unit, the individual chlorophyll-binding sites in each LHC-II monomer are occupied by one type of chlorophyll (either Chla or Chlb). No mixed binding sites were observed. All central ligands of the 14 chlorophylls have been identified as side chains of seven amino-acid residues, two backbone carbonyls, four water molecules and the phosphodiester group of a PG (Supplementary Table 1; a comparison with a previous model⁷ is also included). This coordination mode of Chla 611 to PG is the second case of its kind since its first discovery in photosystem I (ref. 10). On the other side, the phosphodiester group of PG forms a hydrogen-bonding and ionic interaction with the side chains of Tyr 44 and Lys 182 respectively (Fig. 1c).

The polypeptide backbone NH and side chains also form hydrogen bonds with the C7-formyl groups (Chlb) and the C13¹-keto groups of several chlorophylls (Supplementary Table 1). These interactions will not only strengthen the linkage between pigments and protein, but also influence the absorption characteristics of chlorophylls as shown previously¹¹. Except for Chlb 601, nearly all Chlb in the complex are selectively hydrogen-bonded to the polypeptide or to the coordinated water of Chlb 607 through their C7-formyls. The amide side chain of Gln 131 interacts with three Chlb molecules. One hydrogen bond is formed through the interaction of its C = O with the coordinated water of Chlb 606, and two additional hydrogen bonds are formed by its NH₂ interacting with the C7-formyls of Chlb 607 and Chlb 609. Moreover, the C7-formyl of Chlb 606 is hydrogen-bonded to the coordinated water of Chlb 607. All these interactions bring three Chlb into close proximity,

resulting in the clustering of Chlb molecules in this region, which may facilitate the efficient energy transfer between these chlorophylls. It was suggested by functional investigations that Gln 131 is involved in the selective binding of Chlb molecules to LHC-II^{12,13}. As for the selective binding of Chla, we notice that the environment surrounding the C7-methyl groups of Chla molecules is mostly nonpolar. Hydrophobic repulsion or steric hindrance may be the factors affecting the binding affinity of Chlb to these Chla-binding sites.

Chlorophyll arrangement for efficient light harvesting

The chlorophylls in LHC-II are vertically distributed into two layers within the membrane, each layer lying close to the stromal or luminal surface (Fig. 4a). Inside a monomer, the layer close to the stromal surface contains eight chlorophylls (five Chla and three Chlb), which surround the central helices A and B more or less evenly to form an elliptical ring (Fig. 4b). The average centre-to-centre distance between two neighbouring chlorophylls is about 11.26 Å, with a maximum of 12.79 Å and a minimum of 9.74 Å. Each chlorophyll inside this layer can find its symmetric mate related by the internal pseudo-C₂ axis. The remaining six chlorophylls (three Chla and three Chlb) are arranged in the layer close to the luminal surface. They form two separate clusters comprising four chlorophylls (three Chlb and one Chla) and a Chla–Chla dimer (Fig. 4c). Among them, Chlb 606 and Chla 604 are associated with the smallest centre-to-centre distance (8.05 Å) in LHC-II. The shortest distance between two chlorophyll layers is about 13.89 Å (Chlb 609 to Chlb 606).

Another interesting feature of this chlorophyll arrangement is the enrichment of Chlb molecules around helix C and at the interface between monomers (Fig. 4a). All six Chlb molecules are located in this region, with five of them belonging to one monomer and the remaining one (Chlb 601) from the neighbouring monomer. Chlb 601 (II) and Chlb 609 (I) (distance, 11.79 Å) are the closest associated couple of chlorophylls between adjacent monomers within a trimeric LHC-II, indicating that this Chlb-rich region is of critical importance in energy equilibrating inside a functional trimer.

In the trimeric LHC-II, all 24 chlorophylls from the stromal layer are organized into two irregular circular rings (Fig. 4d). The inner ring located in the core region of a trimer is composed of six Chla molecules that are thought to have an important role in inter-monomeric energy transfer¹⁴. The remaining nine Chla and nine Chlb (those covered by the yellow circular ring in Fig. 4d) form the outer ring and are arranged in a mosaic pattern, with three Chlb alternating with three Chla. This new pigment arrangement would favour the efficient absorption of incident light energy from all directions in a broad spectral region and the transfer of the excitation energy to the nearest exit, the putative terminal fluorescence emitter Chla 612 (Supplementary Table 2), in a few steps and at high rates. Energy transfer between two luminal clusters are much less efficient than those within a stromal layer, as they are separated by larger distances (Fig. 4e). We infer that these luminal chlorophyll clusters might serve as upstream energy collectors, absorbing energy and transmitting it to the stromal chlorophylls in a relatively independent way. The energy absorbed by the stromal chlorophylls is quickly focused on Chla 612/Chla 611 and is further transmitted to the neighbouring LHCs or reaction centres.

Carotenoids as light-harvesting antennae

The two central carotenoids with all-*trans* configurations are bound in the grooves on both sides of the supercoil (helices A and B) to form a cross-brace. They are assigned as lutein molecules (Fig. 4). Best fit with the electron density is achieved when the β-rings of both lutein molecules are oriented towards the luminal surface and the ε-rings point to the stromal surface. The polyene chains of lutein 620 and lutein 621 are inclined with respect to the membrane normal

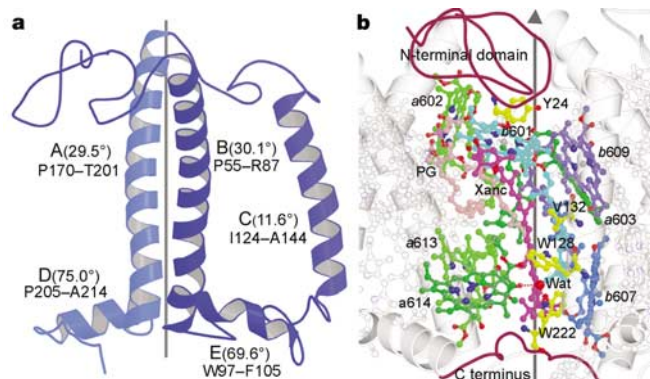


Figure 3 Secondary structure of monomeric LHC-II apoprotein and trimerization. View in parallel with the membrane plane. **a**, The vertical line indicates the approximate direction of the membrane normal and the position of the pseudo-C₂ axis. Helices are labelled A–E. Helix E is newly defined, whereas others are labelled as before⁷. The angle between the central axis of each helix and the membrane normal is shown in parentheses, with the residue range marked below each value. **b**, The interface between two adjacent monomers is shown. Colour code: yellow, amino-acid residues; green, Chla; cyan and blue, Chlb; magenta, xanthophyll-cycle carotenoids; pink, PG; red, water; maroon, C_α traces of N-terminal (Ser 14–Asp 54) and C-terminal (Asp 215–Gly 231) polypeptide chain. The vertical line represents the local C₃ axis of an LHC-II trimer.

by angles of about 59° and 62° , respectively. Both ring-shaped end groups of these two lutein molecules interact with four internal homologous segments of the polypeptide¹⁵ located on both ends of helices A and B through van der Waals contacts and hydrogen bonds. Their polyene chains are firmly fixed in two elongated narrow hydrophobic cavities on both sides of the supercoil, providing strong and rigid linkage between helices A and B. They are indispensable for proper *in vitro* folding of LHC-II into stable complexes^{16–19}.

The third carotenoid, shaped like a bent-over hook, is located in the Chlb-rich region around helix C and is assigned as 9'-*cis* neoxanthin (Fig. 4). Its polyene chain forms an angle of about 58° with the membrane normal. A value of about $57 \pm 1.5^\circ$ derived from linear dichroism spectra²⁰ confirms our assignment. The epoxycyclohexane ring of neoxanthin hangs over the chlorin ring of Chla 604 and is hydrogen-bonded to the hydroxyl of Tyr 112 via its C3'-hydroxyl. Side chains of Leu 134, Met 135, Val 138 from helix C and Trp 71 from helix B as well as chlorin rings and phytol chains of Chlb 606 and Chlb 608 form a hydrophobic cleft that accommodates the hook-shaped polyene chain of neoxanthin. This binding site has been shown to be highly selective for neoxanthin^{17,19}. The cyclohexane ring of neoxanthin on the other end stretches into the exterior solvent region and exhibits weak electron density.

The rate of singlet excitation energy transfer between carotenoids and chlorophylls is correlated with the mutual orientation between them, the centre-to-centre intermolecular distance and the closest distance between two conjugated parts (Supplementary Table 3). Six Chla are found to be in favourable orientations and distances with respect to two luteins for efficient singlet energy transfer from lutein to Chla. The data also show that efficient energy transfer from

neoxanthin to Chlb 606 and Chlb 608 is highly possible. There is experimental evidence to suggest that singlet excitation energy of luteins is transferred exclusively to Chla molecules and not to Chlb²¹. Neoxanthin was found to transfer its energy mostly towards Chlb^{21,22}. It can be concluded that lutein and neoxanthin found in LHC-II may function as effective accessory light-harvesting antennae, absorbing light in the blue–green spectral region as a complement to Chla/b absorbing in the red region. This is in addition to their obvious structural role as well as their photoprotective role of quenching triplet chlorophylls and singlet oxygen⁷.

Structure-based non-photochemical quenching model

The fourth carotenoid we discovered in LHC-II is located at the monomer–monomer interface. The polyene chain of this carotenoid has an all-*trans* configuration and forms a small angle (34°) with the membrane normal. As shown in Fig. 3b, a hydrophobic pocket is formed at the interface by several chlorophylls, hydrophobic residues from the polypeptide and the PG. Part of the polyene chain of this carotenoid together with one of its end groups is accommodated inside this pocket. The opposite end group sticks outside the binding pocket and faces the chlorin plane of Chlb 601 at the stromal side. The two ring-shaped end groups of this carotenoid exhibit distinct electron densities (one flat and the other bulgy between C-5 and C-6). This observation led to our original assignment of this carotenoid as an antheraxanthin, an intermediate in the xanthophyll cycle. However, later carotenoid composition analysis revealed that the major component of xanthophyll-cycle carotenoids in the LHC-II preparation used for crystallization is violaxanthin. We propose that the electron density may be interpreted by a mixed binding of different xanthophyll-cycle carotenoids at this site. The end group at the luminal side points to the cavity formed

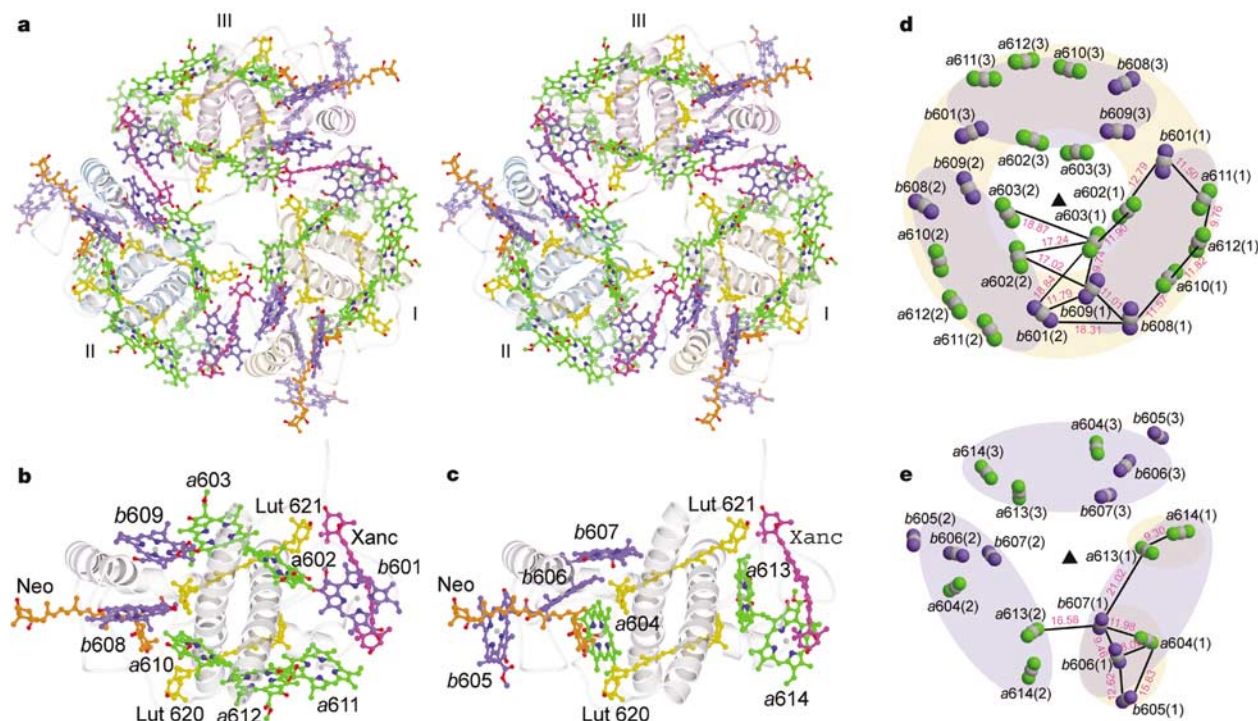


Figure 4 Pigments in the LHC-II trimer and monomer. **a**, Stereo view showing the pigment arrangement pattern in the LHC-II trimer. View along the membrane normal from the stromal side. Monomers are labelled I–III. For clarity, the chlorophyll phytol chains and lipids are omitted. Green, Chla; blue, Chlb; yellow, lutein; orange, neoxanthin; magenta, xanthophyll-cycle carotenoids. **b**, **c**, Pigment pattern in a monomer at the stromal and luminal sides, respectively. Colour designation the same as in **a**. **d**, **e**, Arrangement of chlorophylls within a LHC-II trimer at the stromal and luminal sides, respectively.

Chlorophylls are represented by three atoms: the central magnesium atom and two nitrogen atoms. The connecting line between the two nitrogens defines the directions of the Q_y transition dipole. Green, Chla nitrogen; blue, Chlb nitrogen; grey, magnesium; purple and blue ellipse, approximate monomer area. The magenta numerical note near the dark line connecting two chlorophylls indicates the centre-to-centre distance (Å) between them.

around the local C_3 axis, suggesting that this cavity might be the docking site for violaxanthin de-epoxidase²³.

The xanthophyll cycle was proposed to have a major role in adjusting the efficiency of light harvesting^{4,24,25}. It involves the conversion of violaxanthin into zeaxanthin through antheraxanthin. It was suggested that zeaxanthin molecules can act as direct quenchers of excess excitation by accepting singlet energy transferred from chlorophyll^{26,27}. We observed that at least three chlorophylls are close to the xanthophyll-cycle carotenoids and adopt favourable orientations for efficient singlet excitation transfer from chlorophylls to the xanthophyll-cycle carotenoids (Supplementary Table 3). In addition, we noticed that the distance between Chla 613 and Chla 614 is smaller than 10 Å, and their mutual orientation is close to an irregular distribution (Fig. 4e; Supplementary Table 2). These features agree well with the characteristics of statistical pair energy trap²⁸. We speculate that this pair of Chla molecules might also function as an excitation energy quencher, which would enhance the quenching effect of the xanthophyll-cycle carotenoids.

Here we propose a structure-based non-photochemical quenching (NPQ) model concerning LHC-II (Fig. 5). Efficient non-photochemical energy-transfer pathways are established upon aggregation of LHC-II trimers mediated by DGDG, so that the excitation energy is able to escape from one trimer to the adjacent trimer via these pathways. The following step is the migration of the excitation to the trapping site (the xanthophyll-cycle carotenoids and/or Chla 613–Chla 614 pair), where the non-radiative dissipation of excitation energy happens. This explains the NPQ observed upon incorporation of LHC-II into the liposomes containing DGDG²⁹. Conformational change induced by the protonation of photosystem II proteins including LHC-II was found to be necessary for NPQ^{4,25}. Among the seven lumen-exposed acidic residues in our structure, four of them (Glu 94, Asp 111, Glu 207 and Asp 211) form ion pairs with basic residues. The protonation of these acidic residues under low luminal pH conditions may trigger the conformational changes of helix D and the BC loop. The linker chlorophylls (Chla 614 and Chlb 605) at the trimer–trimer interface (Fig. 5) are coordinated to these regions of polypeptide chain. They may be moved and reoriented to promote the non-photochemical energy transfer and/or the quenching effect of the putative

energy-trapping sites. As a consequence, the potential damaging effect of excess energy would be avoided. □

Methods

The LHC-II was isolated according to the protocol described previously³⁰. A single step of gel filtration chromatography with Hiload 16/60 Superdex 200 pg column (Pharmacia Biotech) was added to improve sample purification, ensuring crystallization reproducibility. The purified LHC-II was solubilized in a solution containing 0.8% *n*-nonyl- β -D-glucoside (BNG) (Anatrace) and 2 mg ml⁻¹ DGDG (Lipid Products) to a final concentration of 4 mg ml⁻¹ chlorophyll (about 8 mg ml⁻¹ protein) and mixed with the crystallization solution containing 66.5 mM HEPES-NaOH pH 7.5, 0.9–1.1 M citrate trisodium and 0.2% *N,N*-bis-(3-D-glucanamidopropyl)deoxycholate (DBC) (Anatrace) in a ratio of 3.0:1.8 (v/v). The resulting drop was equilibrated against a well of 1 ml crystallization solution at 291–293 K using the sitting-drop vapour-diffusion method. Green tabular crystals appeared a week later and grew to a maximum size of about 0.5 × 0.5 × 0.05 mm after one month. Heavy-atom derivative was prepared by soaking the crystal for about 24 h in artificial mother liquor (50 mM HEPES-NaOH pH 7.5, 0.6% BNG, 0.1% DBC, 1.5 mg ml⁻¹ DGDG, 1.09 M citrate trisodium) containing 0.5 mM K₂HgI₄, followed by a backsoaking procedure for about 3–5 h in heavy-atom-free mother liquor. A cryoprotectant (50 mM HEPES-NaOH pH 7.5, 0.4% BNG, 0.15% DBC, 1.0 mg ml⁻¹ DGDG, 1.15 M citrate trisodium, 11% saturated sucrose) was introduced to the crystal by soaking for a few minutes and then the crystal was flash-frozen for the X-ray diffraction experiment. The first native data set and the derivative data set were collected at PF (Tsukuba, Japan) beamline BL6B and the second native data set was collected at BSRF (Beijing, China) beamline 3W1A. A large crystal-to-detector distance (250–350 mm) and a small oscillation (0.5°) are necessary for reducing the overlap of reflections in the large diffraction angle region. Data were processed with Denzo and Scalepack³¹. A typical icosahedral '532' point group symmetry was found by calculating self-rotation function with GLRF³². We inferred that there is only one $T = 1$ icosahedral particle residing in a primitive rhombohedral unit cell by analysing the crystal packing. The icosahedron is oriented with one of its '32' subgroups superposing with the crystallographic '32' point group. Each crystallographic asymmetric unit contains one-sixth of the icosahedron.

Ten heavy-atom sites were located using SnB³³ in direct method mode at 5 Å. The arrangement of heavy atoms in the unit cell also abides with the icosahedral '532' symmetry, confirming our judgement based on self-rotation function and crystal packing analysis. The initial SIR phase was calculated with MLPHARE of the CCP4 suite³⁴ at 5 Å. Phase refinement and extension was performed according to the standard molecular replacement real-space averaging protocol⁹. The NCS matrix was derived from the output of GLRF and improved by IMP³⁵. A molecular envelope enclosed by two icosahedral C_5 axes and one icosahedral C_3 axis covering an icosahedral asymmetric unit with a thickness of about 50 Å was generated by MAMA³⁵. Electron density was calculated using FFT (CCP4) and density averaging was performed with AVE³⁵. The SFALL program (CCP4) was used to calculate F_c and α_c from the averaged density map. The correlation coefficient and R factor were calculated using Rstas (CCP4). The α_c were combined with the initial SIR phases α_b using SigmaA (CCP4) and a new electron-density map was calculated with the F_o and the combined phases. Phases were extended from 5.0 Å to 3.5 Å using a step size of 0.05 Å, and then to 2.72 Å with 0.02 Å step size. Phase refinement was performed for ten iterative cycles in each extension step. The final averaged electron-density map was of high quality, and the backbone was traced without much difficulty. Most of the pigments are clearly defined in the map. The electron-density map was interpreted using the program O³⁶. The polypeptide model was built with the help of a published primary sequence of Lhcb1 (ref. 37). The crystals mainly contain two highly homologous polypeptides, Lhcb1 and Lhcb2. Differences between the two polypeptides are confined to the N-terminal region, which may account for the weakness of the electron density in the region before Ser 14.

Structure refinement was performed with CNS³⁸. At initial stages, two rounds of NCS-constrained simulated annealing (torsion angle dynamics protocol) were performed at 3.5 Å, followed by positional refinement. After the resolution was extended to 2.72 Å, the NCS constraint was switched to restraint mode and the restraint was gradually released at final stages. The individual B factors were refined using 10–2.72 Å data. Peaks above $3.0 \times \sigma$ in the $F_o - F_c$ electron density map were picked as candidates for water molecules, after the R factor was reduced to below 23%. Stereochemical restraints were introduced between chlorophylls and their central ligands during refinement. The final model was evaluated with PROCHECK³⁹ and all ten monomers have good geometry with only one Ramachandran outlier (Val 119) per monomer, the backbone carbonyl of which coordinates a chlorophyll. Secondary structures of polypeptides were analysed with STRIDE⁴⁰. Figures 1c, d and 2b was prepared with program O³⁶. All other figures were prepared with Molscript⁴¹ and Raster3d⁴².

Received 29 October 2003; accepted 27 January 2004; doi:10.1038/nature02373.

- Peter, G. F. & Thornber, J. P. Biochemical composition and organization of higher plant photosystem II light-harvesting pigment-proteins. *J. Biol. Chem.* **266**, 16745–16754 (1991).
- Ruban, A. V., Lee, P. J., Wentworth, M., Young, A. J. & Horton, P. Determination of the stoichiometry and strength of binding of xanthophylls to the photosystem II light harvesting complexes. *J. Biol. Chem.* **274**, 10458–10465 (1999).
- Nußberger, S., Dörner, K., Wang, D. N. & Kühlbrandt, W. Lipid–protein interactions in crystals of plant light-harvesting complex. *J. Mol. Biol.* **234**, 347–356 (1993).
- Horton, P., Ruban, A. V. & Walters, R. G. Regulation of light harvesting in green plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 655–684 (1996).
- Elrad, D., Niyogi, K. K. & Grossman, A. R. A major light-harvesting polypeptide of photosystem II functions in thermal dissipation. *Plant Cell* **14**, 1801–1816 (2002).
- Nilsson, A. *et al.* Phosphorylation controls the three-dimensional structure of plant light harvesting

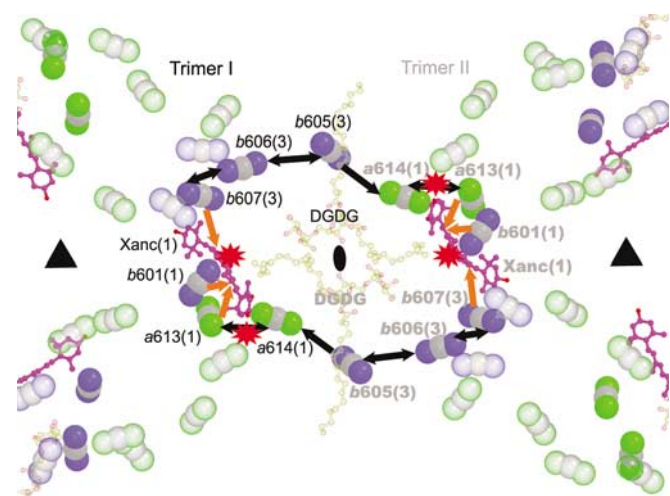


Figure 5 Structure-based non-photochemical chlorophyll fluorescence quenching model in oligomeric LHC-II. Top view along the icosahedral C_2 axis from the stromal side. DGDG is shown as a yellow transparent ball-and-stick model. Chlorophylls and xanthophyll-cycle carotenoids are represented as in previous figures. Black arrows represent the excitation energy-transfer pathways from one trimer to the neighbouring trimer and the orange arrow shows the possible transfer pathways from chlorophyll Q_y to xanthophyll-cycle carotenoids S_1 . The red stars indicate the putative quenching sites. For clarity, characters in one trimer are in black and those of the other are in grey.

- p complex II.
- J. Biol. Chem.*
- 272**
- , 18350–18357 (1997).
7. Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. Atomic model of plant light-harvesting complex by electron crystallography. *Nature* **367**, 614–621 (1994).
 8. Michel, H. Crystallization of membrane proteins. *Trends Biochem. Sci.* **8**, 56–59 (1983).
 9. Rossmann, M. G. in *Methods in Macromolecular Crystallography* (eds Turk, D. & Johnson, L.) 95–103 (IOS Press, Amsterdam, 2001).
 10. Jordan, P. *et al.* Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411**, 909–917 (2001).
 11. McLuskey, K., Prince, S. M., Cogdell, R. J. & Isaacs, N. W. The crystallographic structure of the B800–820 LH3 light-harvesting complex from the purple bacteria *Rhodospseudomonas acidophila* strain 7050. *Biochemistry* **40**, 8783–8789 (2001).
 12. Bassi, R., Croce, R., Cugini, D. & Sandonà, D. Mutational analysis of a higher plant antenna protein provides identification of chromophores bound into multiple sites. *Proc. Natl Acad. Sci. USA* **96**, 10056–10061 (1999).
 13. Remelli, R., Varotto, C., Sandonà, D., Croce, R. & Bassi, R. Chlorophyll binding to monomeric light-harvesting complex. A mutation analysis of chromophore-binding residues. *J. Biol. Chem.* **274**, 33510–33521 (1999).
 14. Gradinaru, C. C. *et al.* The flow of excitation energy in LHClI monomers: implications for the structural model of the major plant antenna. *Biophys. J.* **75**, 3064–3077 (1998).
 15. Bassi, R., Sandonà, D. & Croce, R. Novel aspects of chlorophyll *a/b*-binding proteins. *Physiol. Planta.* **100**, 769–779 (1997).
 16. Plumley, F. G. & Schmidt, G. W. Reconstitution of chlorophyll *a/b* light-harvesting complexes: Xanthophyll-dependent assembly and energy transfer. *Proc. Natl Acad. Sci. USA* **84**, 146–150 (1987).
 17. Croce, R., Weiss, S. & Bassi, R. Carotenoid-binding sites of the major light-harvesting complex II of higher plants. *J. Biol. Chem.* **274**, 29613–29623 (1999).
 18. Paulsen, H., Finkenzeller, B. & Kuhllein, N. Pigments induce folding of light-harvesting chlorophyll *a/b*-binding protein. *Eur. J. Biochem.* **215**, 809–816 (1993).
 19. Hobe, S., Niemeier, H., Bender, A. & Paulsen, H. Carotenoid binding sites in LHClIb. Relative affinities towards major xanthophylls of higher plants. *Eur. J. Biochem.* **267**, 616–624 (2000).
 20. Croce, R., Remelli, R., Varotto, C., Breton, J. & Bassi, R. The neoxanthin binding site of the major light harvesting complex (LHClI) from higher plants. *FEBS Lett.* **456**, 1–6 (1999).
 21. Gradinaru, C. C., Stokkum, I. H. M. V., Pascal, A. A., Grondelle, R. V. & Amerongen, H. V. Identifying the pathways of energy transfer between carotenoids and chlorophylls in LHClI and CP29. A multicolor, femtosecond pump-probe study. *J. Phys. Chem. B* **104**, 9330–9342 (2000).
 22. Croce, R., Müller, M. G., Bassi, R. & Holzwarth, A. R. Carotenoid-to-chlorophyll energy transfer in recombinant major light-harvesting complex (LHC-II) of higher plants. I. femtosecond transient absorption measurements. *Biophys. J.* **80**, 901–915 (2001).
 23. Hieber, A. D., Bugos, R. C. & Yamamoto, H. Y. Plant lipocalins: violaxanthin de-epoxidase and zeaxanthin epoxidase. *Biochim. Biophys. Acta* **1482**, 84–91 (2000).
 24. Demmig-Adams, B. & Adams, W. W. Photoprotection and other responses of plants to high light stress. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**, 599–626 (1992).
 25. Gilmore, A. M. Mechanistic aspects of xanthophyll cycle-dependent photoprotection in higher plant chloroplasts and leaves. *Physiol. Planta.* **99**, 197–209 (1997).
 26. Frank, H. A. *et al.* Photophysics of the carotenoids associated with the xanthophyll cycle in photosynthesis. *Photosynth. Res.* **41**, 389–395 (1994).
 27. Ma, Y. Z., Holt, N. E., Li, X. P., Niyogi, K. K. & Fleming, G. R. Evidence for direct carotenoid involvement in the regulation of photosynthetic light harvesting. *Proc. Natl Acad. Sci. USA* **100**, 4377–4382 (2003).
 28. Beddard, G. S., Carlin, S. E. & Porter, G. Concentration quenching of chlorophyll fluorescence in bilayer lipid vesicles and liposomes. *Chem. Phys. Lett.* **43**, 27–32 (1976).
 29. Moya, I., Silvestri, M., Vallon, O., Cinque, G. & Bassi, R. Time-resolved fluorescence analysis of the photosystem II antenna proteins in detergent micelles and liposomes. *Biochemistry* **40**, 12552–12561 (2001).
 30. Lou, S., Wang, K., Zhao, F., Xu, C. & Kuang, T. A comparative study on PS II light harvesting chlorophyll *a/b* protein complexes between spinach and cucumber. *Acta Bot. Sin.* **37**, 192–197 (1995).
 31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
 32. Tong, L. & Rossmann, M. G. The locked rotation function. *Acta Crystallogr. A* **46**, 783–792 (1990).
 33. Weeks, C. M. & Miller, R. The design and implementation of SnB v2.0. *J. Appl. Crystallogr.* **32**, 120–124 (1999).
 34. CCP4 Collaborative Computational Project, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
 35. Jones, T. A. in *Molecular Replacement* (eds Dodson, E. J., Gover, S. & Wolf, W.) 92–105 (SERC Daresbury Laboratory, Warrington, 1992).
 36. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
 37. Mason, J. G. Nucleotide sequence of a cDNA encoding the light-harvesting chlorophyll *a/b* binding protein from spinach. *Nucleic Acids Res.* **17**, 5387 (1989).
 38. Brünger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
 39. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
 40. Frishman, D. & Argos, P. Knowledge-based secondary structure assignment. *Proteins Struct. Funct. Genet.* **23**, 566–579 (1995).
 41. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
 42. Merritt, E. A. & Murphy, M. E. P. Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. C. Liang and the late P. S. Tang for their efforts in initiating this project; X. C. Gu for discussions; N. Sakabe and K. Sakabe at PF (Tsukuba, Japan) and the staff at BSRF (Beijing, China) for their support during data collection at the synchrotron facilities. This research was financially supported by the National Key Research Development Project of China, the National Natural Science Foundation of China, the National Key Special Research Program and the Knowledge Innovation Program of the Chinese Academy of Sciences.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.R.C. (wrchang@sun5.ibp.ac.cn). Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 1RW7.

Compact sources as the origin of the soft γ -ray emission of the Milky Way

F. Lebrun¹, R. Terrier^{1,2}, A. Bazzano³, G. Bélanger¹, A. Bird⁴, L. Bouchet⁵, A. Dean⁴, M. Del Santo³, A. Goldwurm¹, N. Lund⁶, H. Morand¹, A. Parmar⁷, J. Paul^{1,2}, J.-P. Roques⁵, V. Schönfelder⁸, A. W. Strong⁸, P. Ubertini³, R. Walter⁹ & C. Winkler⁷

¹CEA-Saclay, DAPNIA/Service d'Astrophysique, F91191 Gif sur Yvette Cedex, France

²Fédération de Recherche APC, Collège de France, 11, place Marcelin Berthelot, F75231 Paris, France

³IASF/CNR, Via del fosso del cavaliere 100, 00133 Roma, Italy

⁴School of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK

⁵CESR, 9 avenue du colonel Roche – BP 4346, 31028 Toulouse Cedex 4, France

⁶DSRI, Julian Maries Vej 30, DK-2100 Copenhagen 0, Denmark

⁷ESA, RSSD, Keplerlaan 1, NL-2201 AZ Noordwijk, The Netherlands

⁸MPE, Giessenbachstrasse, 85748 Garching bei München, Germany

⁹ISDC, Chemin d'Ecogia, CH-1290 Versoix, Switzerland

The Milky Way is known to be an abundant source of γ -ray photons¹, now determined to be mainly diffuse in nature and resulting from interstellar processes². In the soft γ -ray domain, point sources are expected to dominate, but the lack of sensitive high-resolution observations did not allow for a clear estimate of the contribution from such sources^{3,4}. Even the best imaging experiment⁵ revealed only a few point sources, accounting for about 50% of the total Galactic flux⁶. Theoretical studies were unable to explain the remaining intense diffuse emission^{7,8}.

Investigating the origin of the soft γ -rays is therefore necessary to determine the dominant particle acceleration processes and to gain insights into the physical and chemical equilibrium of the interstellar medium⁷. Here we report observations in the soft γ -ray domain that reveal numerous compact sources. We show that these sources account for the entirety of the Milky Way's emission in soft γ -rays, leaving at most a minor role for diffuse processes.

There are two main processes that can lead to an interstellar soft γ -ray emission. The first and the most natural is inverse Compton scattering of high-energy (GeV) cosmic-ray electrons on the ambient photon field. However, the electrons required to account for the bulk of the Galactic emission would also produce radio-synchrotron emission in the Galactic magnetic field at a level much higher than the one actually observed⁹. The second assumes the presence of a population of electrons of a few hundred keV radiating through bremsstrahlung interactions with the interstellar gas. Because these electrons would lose their energy primarily through ionization and Coulomb collisions, the total power required to compensate for their energy losses is of the order of 10^{41} – 10^{43} erg s⁻¹ (ref. 10). This power, comparable or higher than that of the cosmic ray protons, would affect the interstellar-medium ionization equilibrium and give rise to an excessive dissociation of the interstellar molecules. In the light of the above arguments outlining the improbability of an interstellar origin for the galactic soft γ -ray emission, and taking into account the spectral shape of this emission, the hypothesis of a major point-source contribution was put forth^{10,11}.

The INTEGRAL γ -ray observatory¹², launched in October 2002, carries two major co-aligned coded-mask instruments: the IBIS imager¹³ and the SPI spectrometer¹⁴. To lessen the instrumental

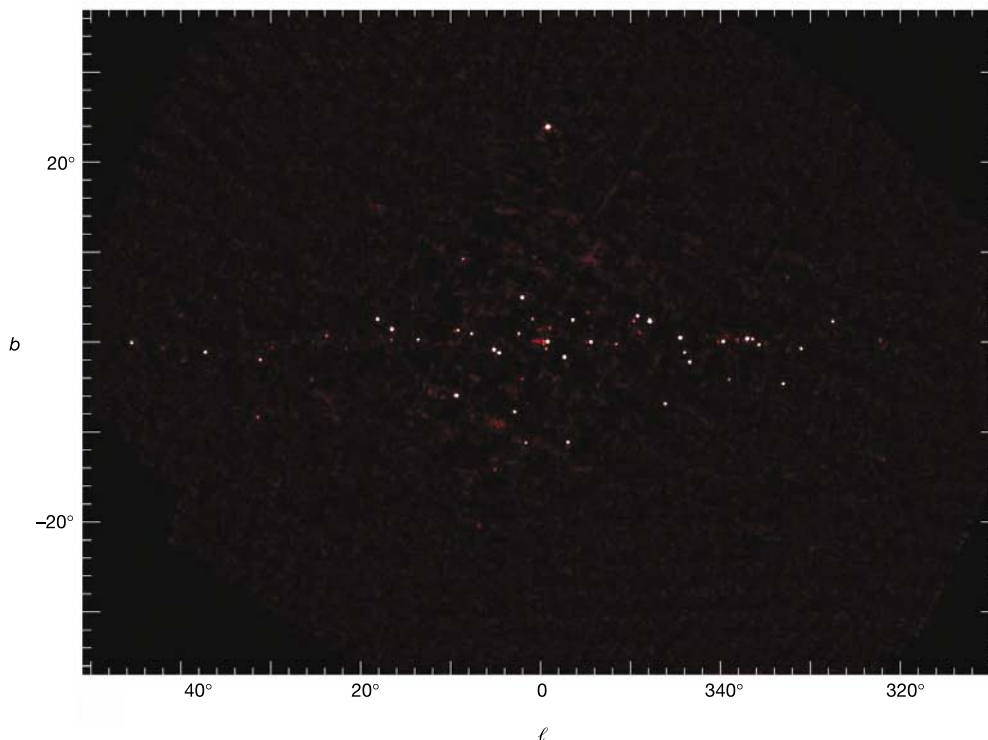


Figure 1 IBIS/ISGRI view of the Galactic Centre region. ISGRI galactic map of the central part of the Milky Way in the 20–60 keV energy band. Most of the 91 sources detected in this image lay in the Galactic plane. The Galactic bulge, the region around the Galactic Centre ($l = 0^\circ$, $b = 0^\circ$), is densely populated, especially with low-mass X-ray binaries. This map is a mosaic of about 2,000 images resulting from the IBIS standard analysis¹⁶ of 2.2-ks pointed observations. The standard analysis deconvolves detector images with the

mask pattern, detects the significant peaks and subtracts the point spread function (PSF) side-lobes in the resulting sky images. ISGRI¹⁵, the low-energy camera of IBIS, uses 16,384 independent CdTe detectors operating between 15 keV and 1 MeV. With the IBIS mask 3.2 m above, it provides a $13'$ angular resolution over a $19^\circ \times 19^\circ$ FWHM field of view and has a sensitivity close to 1 mCrab below 100 keV for a 10^6 -s observation. The average point-source location accuracy is a few arcmin²⁰.

background contribution, both experiments are actively shielded. Coded mask imaging acts as a high-pass spatial-frequency filter, strongly attenuating structures larger than the instrument's angular resolution, which is $13'$ for IBIS and 2.6° for SPI. This makes IBIS, with its ISGRI¹⁵ photon-counting γ -camera, very well suited to measure the emission from compact sources. Below 100 keV, the IBIS/ISGRI sensitivity reaches the millicrab level.

INTEGRAL observed the Galactic Centre region in the spring and fall of 2003 as part of its core programme (guaranteed time) and during two target-of-opportunity observations. The total observing time amounts to about 1.5×10^6 s in the central part of the survey. Figure 1 shows an IBIS/ISGRI mosaic image produced using the standard analysis software¹⁶. Sources were detected and removed one by one from the most to the least significant, yielding a list of 91 excesses above 6 sigma. Catalogued counterparts were searched for each of these sources using SIMBAD at the Centre des Données Stellaires de Strasbourg. All but 26 were identified with known sources. Forty of the identified sources are accreting binary systems with low-mass companions as expected in a region that is rich in old stars such as the Galactic bulge.

The remaining sources of known type are: seven high-mass systems, two radio pulsars, two plerionic supernova remnants, one millisecond pulsar, one soft γ -ray repeater and one Seyfert 1 galaxy. This leaves 11 sources of unknown type. The known binaries contribute to the total source flux at levels of 86%, 78%, 77% and 74% respectively in the 20–40, 40–60, 60–120 and 120–220 keV bands. The global spectrum of the other sources is thus significantly harder, possibly indicating a new emerging population of hard sources. Some of the unidentified sources are highly absorbed such as IGRJ16318–4848 (ref. 17), discovered during the course of this survey. A significant contribution from plerions and pulsars could also be considered, as their spectra are much harder than the average.

Diffuse emission is washed out during the standard imaging process, so its contribution can be estimated by comparing the data before and after this processing. This entails comparing the ISGRI count rate and the combined intensity of detected sources. The count rate is due to the internal background, the cosmic diffuse background, the galactic diffuse emission and point sources. If the ISGRI count rates can be corrected for the internal background variations and the contribution from the known point sources removed, we should be left with a Galactic emission, possibly diffuse, and an isotropic contribution; the latter composed of the cosmic diffuse background and the constant internal background.

The internal background has a hard spectrum and fully dominates the count rates above 500 keV. We can therefore use the high-energy ISGRI count-rate to predict the variable part of the internal background. The relation between the count rate in each energy range and that above 500 keV can be derived from high-latitude observations devoid of galactic emission. The variable part of the internal background has been subtracted from each observation window to produce corrected count rates. The contribution of each detected source was computed by applying the IBIS angular response to the intensity at the source position in the field of view. To convert intensities to source count rates, a conversion factor was derived from Crab nebula observations.

Galactic maps of the difference between the corrected and the source count rates were built. Each sky bin was assigned the average of the count rates of all observations pointing within it. The resulting map has a resolution of the order of the field of view (19° full-width at half-maximum, FWHM). These residuals contain an isotropic emission and a Galactic contribution. Their respective importance was fitted to the latitude profile of the residuals. The fitted count rate of this residual Galactic emission is 8.6 ± 3.0 s⁻¹, 0.2 ± 1.1 s⁻¹, 0.7 ± 1.3 s⁻¹, and -0.7 ± 0.6 s⁻¹, respectively in the 20–40, 40–60, 60–120 and 120–220 keV bands. A Galactic diffuse

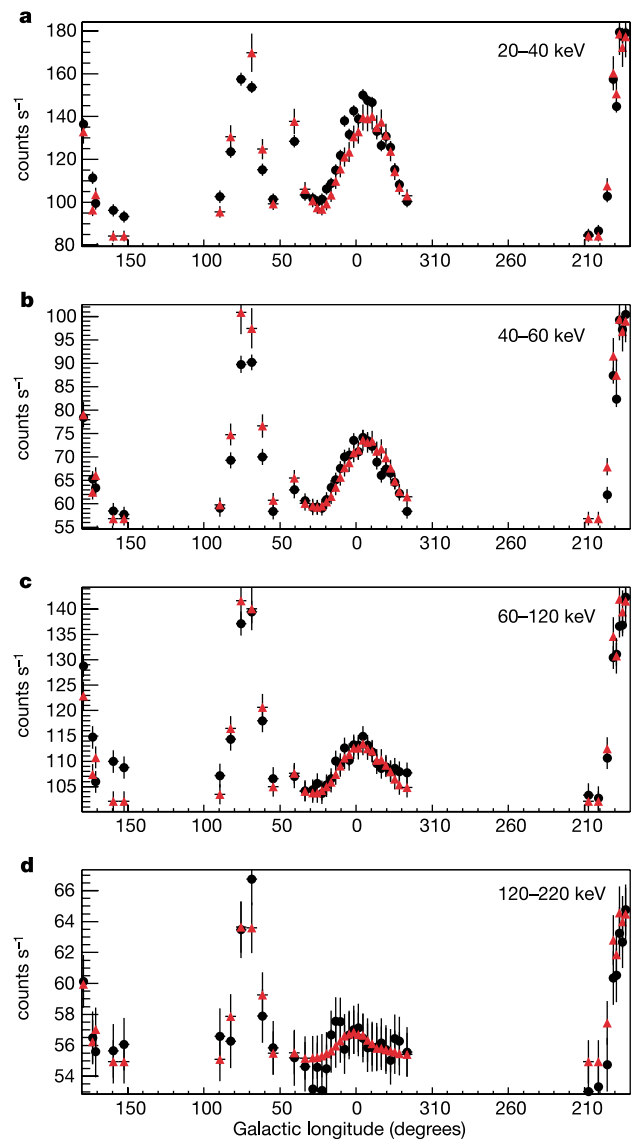


Figure 2 Longitude profile of the Galactic emission at $|b| < 5^\circ$. Comparison between the background-corrected detector count rate (black circles) and the estimated count rate from the detected sources (red triangles) as a function of longitude for each energy band (a–d). These profiles were produced averaging the corrected count rates of all pointings lying in the same longitude bin, with $|b| < 5^\circ$. The resulting angular resolution is then given by the instrument's field of view. Because the variable internal background dominates all other components above 500 keV, the count rate above that energy can be used to predict and subtract the internal background in each energy range. This correction has been calibrated using the relation between the count rates in each energy range and in the range above 500 keV during high-latitude observations (pointing at $|b| > 30^\circ$) that are deprived of Galactic and point-source emission. The contribution of the Galactic emission above 500 keV does not significantly bias the background subtraction. Even if the total Galactic flux above 500 keV were at the Crab nebula level, the bias introduced in the correction would never exceed 0.3 s⁻¹, which is negligible, at least below 120 keV. The total point-source count rate profile is estimated for each pointing correcting for acceptance and attenuation effects due to the source position in the field of view. The flux obtained through the imagery is then calibrated on the background-corrected crab count rate. The error bars are due to the systematic uncertainties introduced during this process. The constant isotropic background obtained by fitting the latitude distribution (see Fig. 3) has been added to the source count rate.

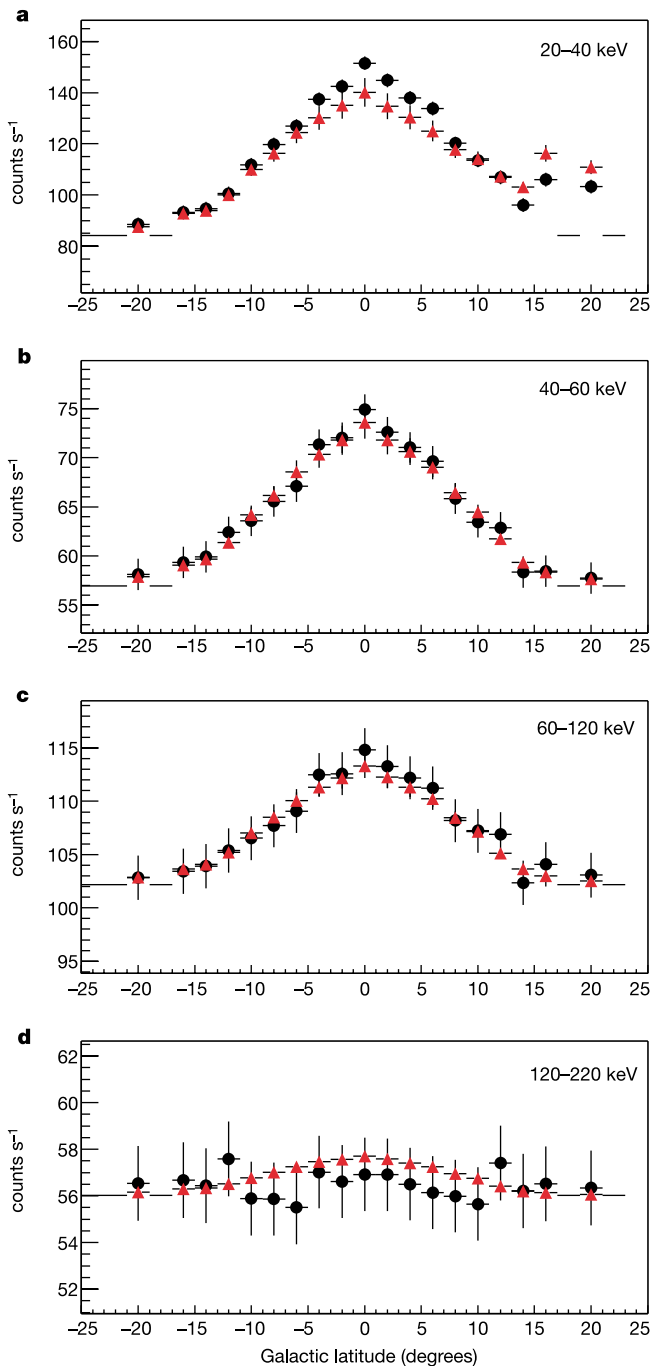


Figure 3 Latitude profiles of the Galactic emission in the central regions where $|l| < 20^\circ$. These profiles (a–d) have been produced as for Fig. 2. They show background-corrected count rates (black circles), and source contribution (red triangles), in the Galactic central regions. The interstellar emission is believed to be distributed as a 5° FWHM gaussian along the Galactic plane³. Such a gaussian has been convolved with the IBIS field of view and the resulting Galactic distribution has a width of around 15° . To estimate the amount of residual emission compatible with Galactic emission, we fitted the latitude profile of the difference between corrected count rate and source count rate, with the previously defined Galactic distribution and an isotropic component. The latter have been added to the source count rate in each profile to see the amount of residual emission. A contribution of 13% is visible under 40 keV (a). Scorpio X-1 ($b \approx 20^\circ$) is not well accounted for because of the inaccuracies of the response model at large angles. No significant emission is seen at higher energies. In the 120–220 keV band, the systematic errors due to the background subtraction and the source count rate normalization limit the significance of the result.

component contribution of 13% of the total Galactic emission appears below 40 keV. There is no indication of an interstellar emission beyond. We note, however, that at high energy (>120 keV) the uncertainty becomes large compared to the total source contribution.

Figure 2 displays the longitude profile of the corrected count rate together with that from the detected sources. The Crab nebula and the equally bright accreting black hole Cygnus X-1 dominate the longitude profiles. At low energy, the Galactic Centre is not much less significant but weakens faster with energy. Although the general agreement between both profiles is very good, particularly at low energies, we notice significant discrepancies. Because our source list is complete only over the central Galactic radian, some of these discrepancies may be due to missing sources. Other discrepancies may be ascribed to inaccurate or inappropriate response functions. In particular, above ~ 100 keV, the passive shield is no longer fully opaque, although we assumed it was in our computation of the source contribution. Latitude profiles of the corrected count rates are given in Fig. 3. The diffuse Galactic contribution at the 13% level revealed in the fitting procedure is clearly apparent on the top profile. This emission could be due to weaker sources and/or interstellar emission. It is peaked both in latitude and longitude and its $\sim 10^\circ$ radius is reminiscent of the Galactic bulge strongly populated with low-mass X-ray binaries. At high energy ($E > 120$ keV), the large systematic uncertainties reflect the limitations of the method. It is assumed that the corrected count rates are only due to emission within the field of view but the passive shield is no longer opaque at these energies.

In this regard, we note that SPI, using a large anticoincidence system, is much better shielded at high energy. Strong *et al.*¹⁸ have made a preliminary attempt to estimate the Galactic diffuse emission, separating only the four brightest sources near the Galactic Centre. We wondered what would have been the effect of the 43 sources detected in the simultaneous ISGRI data. Taking advantage of the summed spectra of the four sources measured by both instruments, a cross-calibration can be extracted. The total Galactic emission measured by SPI can be obtained by summing the ‘diffuse’ and source spectra given in Fig. 2 of ref. 19. The sources detected in the ISGRI data contribute to 86% of the Galactic emission observed with SPI between 100 and 200 keV, giving a heuristic estimate of the source contribution in this energy range.

It appears that the Galactic soft γ -ray emission is fully dominated by compact sources even if a little room still exists for interstellar emission. The difference with the X-ray domain (<10 keV) is particularly remarkable¹⁹. It seems that between 10 and 20 keV a complete change in the mechanisms at the origin of the Galactic emission must occur. However, we should bear in mind the vast difference in spatial resolution between the X-ray and γ -ray experiments. A feature a few arcmin wide will appear as diffuse to Chandra and as a point source to IBIS. Below 40 keV the remaining 13% not accounted for by the detected compact sources could be due to weak bulge sources or to the same mechanism at work below 10 keV. Interstellar processes can still contribute to the soft γ -ray emission of the Galaxy but at such a low level or in so restricted areas that the power supply, ionization or molecule dissociation problems are alleviated. $\square$

Received 8 October 2003; accepted 10 February 2004; doi:10.1038/nature02407.

1. Schönfelder, V. *The Universe in Gamma Rays* (A&A Library, Springer, 2001).
2. Strong, A. W., Moskalenko, I. V. & Reimer, O. Diffuse continuum gamma rays from the Galaxy. *Astrophys. J.* **537**, 763–784 (2000).
3. Kinzer, R. L., Purcell, W. R. & Kurfess, J. D. Gamma-ray emission from the inner Galactic ridge. *Astrophys. J.* **515**, 215–225 (1999).
4. Valinia, A., Kinzer, R. L. & Marshall, F. E. Measurement of the Galactic X-ray/gamma-ray background radiation: contribution of point sources. *Astrophys. J.* **534**, 277–282 (2000).
5. Paul, J. *et al.* SIGMA: the hard X-ray and soft gamma-ray telescope on board the GRANAT space observatory. *Adv. Space Res.* **11**, (8)289–(8)302 (1991).
6. Purcell, W. R. *et al.* Diffuse galactic gamma-ray continuum. *Astron. Astrophys. Suppl. Ser.* **120**, 389–392 (1996).

7. Pohl, M. Limits for an inverse bremsstrahlung origin of the diffuse Galactic soft gamma-ray emission. *Astron. Astrophys.* **339**, 587–590 (1998).
8. Dogiel, V. A., Inoue, H., Schönfelder, V. & Strong, A. W. The origin of diffuse X-ray emission from the Galactic ridge. I. Energy output of particle sources. *Astrophys. J.* **581**, 1061–1070 (2002).
9. Webber, W. R. in *Composition and Origin of Cosmic Rays* (ed. Shapiro, M. M.) 83–100 (Boston, Dordrecht, 1983).
10. Skibo, J. G., Ramaty, R. & Purcell, W. R. Implications of the diffuse Galactic continuum. *Astron. Astrophys. Suppl. Ser.* **120**, 403–406 (1996).
11. Lebrun, F. *et al.* Nature of the Galactic soft gamma-ray emission. *Astrophys. Lett. Commun.* **38**, 457–460 (1999).
12. Winkler, C. *et al.* The INTEGRAL mission. *Astron. Astrophys.* **411**, L1–L6 (2003).
13. Ubertini, P. *et al.* IBIS: the imager on-board INTEGRAL. *Astron. Astrophys.* **411**, L131–L139 (2003).
14. Vedrenne, G. *et al.* The spectrometer aboard INTEGRAL. *Astron. Astrophys.* **411**, L63–L70 (2003).
15. Lebrun, F. *et al.* ISGRI: the INTEGRAL Soft Gamma-Ray Imager. *Astron. Astrophys.* **411**, L141–L148 (2003).
16. Goldwurm, A. *et al.* The INTEGRAL/IBIS scientific data analysis. *Astron. Astrophys.* **411**, L223–L229 (2003).
17. Walter, R. *et al.* INTEGRAL discovery of a bright highly obscured galactic X-ray binary source IGR J16318–4848. *Astron. Astrophys.* **411**, L427–L432 (2003).
18. Strong, A. W. *et al.* Diffuse continuum emission from the inner Galaxy: first results from INTEGRAL/SPI. *Astron. Astrophys.* **411**, L447–L450 (2003).
19. Wang, Q. D., Gotthelf, E. V. & Lang, C. C. A faint discrete source origin for the highly ionized iron emission from the Galactic centre region. *Nature* **415**, 148–150 (2002).
20. Gros, A. *et al.* The INTEGRAL IBIS/ISGRI point spread function and source location accuracy. *Astron. Astrophys.* **411**, L179–L183 (2003).

Acknowledgements This paper is based on observations with INTEGRAL, an ESA project with instruments and science data centre funded by ESA member states (especially the PI countries: Denmark, France, Germany, Italy, Switzerland and Spain), the Czech Republic and Poland, and with the participation of Russia and the USA. We thank CNES for its support during the ISGRI development and in the INTEGRAL data analysis.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.L. (flebrun@cea.fr).

Structural relaxation in supercooled water by time-resolved spectroscopy

Renato Torre^{1,2,3}, Paolo Bartolini^{1,2} & Roberto Righini^{1,2,4}

¹European Laboratory for Non-linear Spectroscopy, ²Istituto Nazionale per la Fisica della Materia, ³Dipartimento di Fisica, and ⁴Dipartimento di Chimica, Università di Firenze, Polo Scientifico, Sesto Fiorentino, Firenze, 50019 Italia

Water has many kinetic and thermodynamic properties that exhibit an anomalous dependence on temperature^{1–5}, in particular in the supercooled phase. These anomalies have long been interpreted in terms of underlying structural causes, and their experimental characterization points to the existence of a singularity at a temperature of about 225 K. Further insights into the nature and origin of this singularity might be gained by completely characterizing the structural relaxation in supercooled water⁶. But until now, such a characterization has only been realized in simulations^{7–9} that agree with the predictions of simple mode-coupling theory¹⁰; unambiguous experimental support for this surprising conclusion is, however, not yet available^{11–14}. Here we report time-resolved optical Kerr effect measurements¹⁵ that unambiguously demonstrate that the structural relaxation of liquid and weakly supercooled water follows the behaviour predicted by simple mode-coupling theory. Our findings thus support the interpretation^{7–9} of the singularity as a purely dynamical transition. That is, the anomalous behaviour of weakly supercooled water can be explained using a fully dynamic model and without needing to invoke a thermodynamic origin. In this regard, water behaves like many other, normal molecular liquids that are fragile glass-formers.

The shear viscosity, self-diffusion coefficient and relaxation times

of water above melting and in the supercooled state all exhibit a temperature dependence following a diverging power law^{1–5} of the type $(T/T_s - 1)^{-x}$, independent of the measurement method used. Thermodynamic properties such as heat capacity, compressibility and thermal expansion coefficient similarly show a temperature dependence characterized by an anomalous increase upon cooling. Different experimental determinations of the singularity temperature T_s converge on values of about 223–228 K. Although the nature and physical origin of this singularity are still debated^{16–18}, the observed anomalous phenomena suggest that modifications of structural properties of water occur around this temperature. Characterizing the correlation functions of water and their time evolution, at temperatures below melting and approaching T_s , might therefore yield new insights. But although simulation results suggest that the features of the investigated dynamic correlation functions agree well with the predictions of mode-coupling theory^{7–9}, unambiguous experimental support for this surprising conclusion is not yet available^{11–14}.

Several recent time-resolved optical Kerr effect (OKE) experiments on water have revealed^{11,19–23} an oscillation on short time-scales (for delay time shorter than 1 ps) in the time profile of the measured signal; at longer times, the signal shows a monotonous

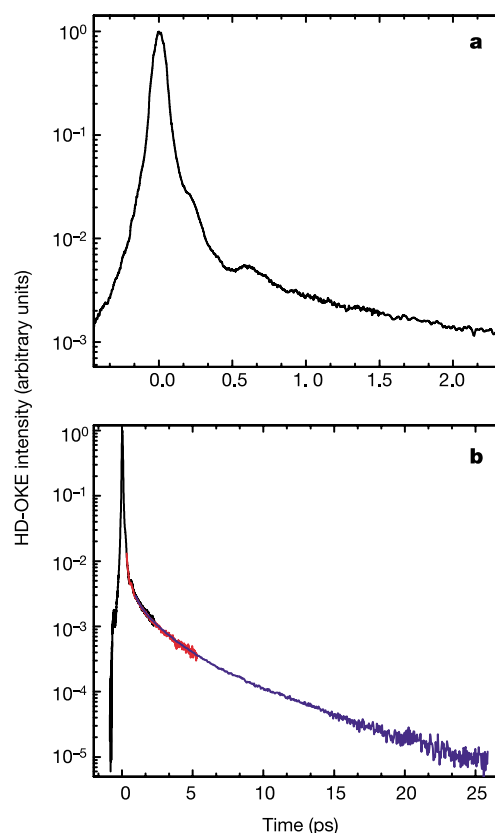


Figure 1 Signal decay measured in the time-resolved HD-OKE experiment on supercooled water at $T = 257$ K. **a**, HD-OKE signal in the short time region. An oscillatory component occurs in the decay trace for time shorter than 1 ps. **b**, HD-OKE signal over the whole time region measured. For longer time the relaxation shows a monotonous decay. The plot illustrates the procedure used for improving the sensitivity. The entire time decay scan was divided into three sections. Short laser pulses (60 fs) were used for the short time scan (black line); for the intermediate time delay intervals (red line) the laser pulse duration was stretched up to 120 fs; for the longer time (blue line) we used stretched laser pulses of 500 fs duration. In this procedure we increased the pulse energy to keep the peak power roughly constant. Care was taken to ensure a large time overlap between subsequent scans, to allow an accurate rescaling of the intensities and a proper reconstruction of the entire decay curve.

decay for which various interpretations have been proposed. The oscillatory part of the signal shows a limited dependence on temperature, while the long time decay slows down as temperature is lowered. So far, attention has mostly focused on the initial oscillatory component of the OKE signal, which provides information about intermolecular hydrogen-bond dynamics and might be directly compared with the low-frequency Raman spectra. But this focus also reflects practical considerations: the intensity of the OKE signal of water is very low, making it very difficult to determine its relaxation profile accurately at times longer than a very few picoseconds.

Here we present the results of an extensive investigation by means of heterodyne detected OKE experiments (HD-OKE) of the relaxation in liquid and supercooled water, using the experimental set-up described in ref. 24. The low noise level and the large dynamic range provide data of sufficiently high quality to extend measurements by almost twenty degrees into the supercooled region. The procedure adopted in our experiments is based on the use of laser pulses of variable duration and energy (see Fig. 1). It allows us to measure the OKE signal relaxation over a large time interval: at the lowest temperature (254 K), we follow the decay of the OKE signal over 40 ps. We can thus provide the first unambiguous measurement of

the entire correlation function in liquid water and perform a full mode-coupling theory (MCT) analysis of the slow dynamic regime.

An important aspect to be considered in interpreting OKE and depolarized light-scattering experiments on liquid water is the microscopic origin of the measured signal. The molecular polarizability of water is almost isotropic and the optically induced polarization is therefore dominated by intermolecular (collision-induced) contributions. This holds not only for the high-frequency ($>10\text{ cm}^{-1}$) part of the light-scattering spectrum (corresponding to the short time OKE signal), but also for the very-low-frequency part²⁵ that corresponds to the long time relaxation tail of our time-domain data. In other words, the OKE and depolarized-light-scattering techniques are mostly sensitive to the intermolecular dynamics of water. In the short delay time limit, this corresponds to the intracage vibrational dynamics; the long time decay, in contrast, is prevalently associated with rearrangement of local (cage) structures, which represents the first step of the overall structural relaxation. More precisely, the measured signal in a OKE experiment is defined by the material response function¹⁵, that is, by the time-derivative of the dielectric-constant correlation function^{26,27} which, in water, is dominated by the intermolecular contribution.

Figure 2a collects several HD-OKE data measured at different temperatures (314–254 K). The log-log plots clearly show the marked slowing down with decreasing temperature of the overall relaxation, and the slight temperature dependence of the fast

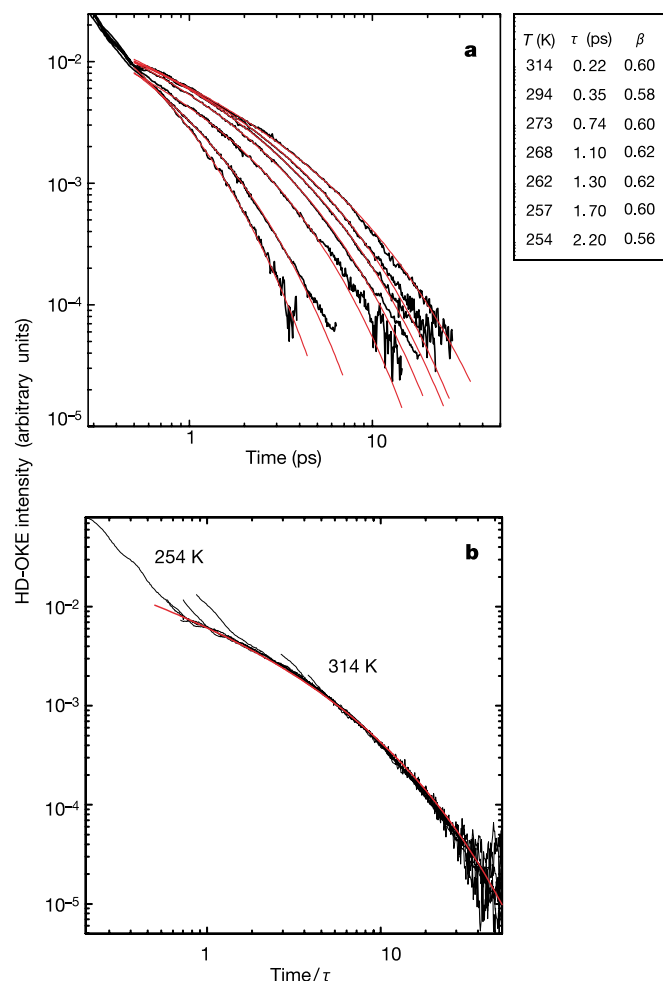


Figure 2 Temperature dependence of the HD-OKE signals on water and fits based on the stretched exponential function. **a**, A log-log plot of the HD-OKE signals as a function of the temperature. The red curves are best fits performed with the time-derivative of the stretched exponential function; the values of the τ and β parameters are listed. **b**, Master plot obtained by rescaling the time and intensity axes of the HD-OKE signals showed in **a**; the reported red line is the master curve corresponding to the time-derivative of the stretched exponential with $\tau = 1$ and $\beta = 0.6$.

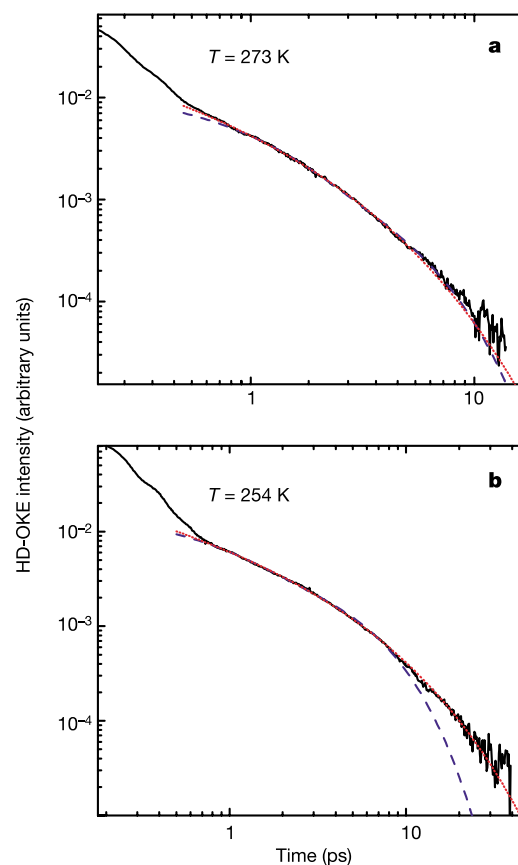


Figure 3 HD-OKE signals on liquid and supercooled water with a detailed analysis of the relaxation time profile. **a**, HD-OKE signal at $T = 273\text{ K}$. The difference between fits performed with the time-derivative of the double exponential (blue dashed line) and time-derivative of the stretched exponential (red dotted line) is hardly observable owing to the noise and time limitations. **b**, HD-OKE signal at $T = 254\text{ K}$. The time-derivative of the double exponential curve (blue dashed line) is clearly unable to reproduce the data for time delays longer than 8 ps—instead the time-derivative of the stretched exponential function is able to fully fit the experimental decay in the complete time region investigated.

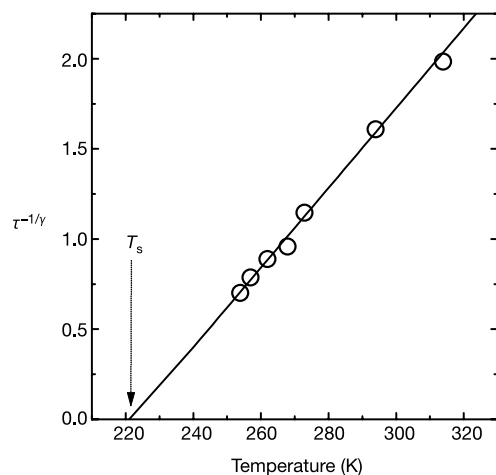


Figure 4 Temperature dependence of the water relaxation times, obtained by the fit of HD-OKE signal decays with the time-derivative of the stretched exponential function. The relaxation times in Fig. 2 follow the power law $\tau(T) = a(T - T_S)^{-\gamma}$ with $T_S = 221 \pm 5$ K and $\gamma = 2.2 \pm 0.3$. Here we reported the data (open circles) and the fit (continuous line) in a linearized plot, $\tau^{-1/\gamma}$ versus T . This plot shows clearly the linear dependence of the renormalized times and the T_S value supported by the fit. The T_S and γ values are in agreement with the MCT, so our data support the equivalence of the singularity temperature T_S with the dynamical critical temperature predicted by the MCT.

component (for $\tau < 1$ ps). Figure 2b shows the master plot obtained from the data of Fig. 2a by rescaling the time and amplitude axes of each curve. The perfect coincidence of the different curves demonstrates that the slow relaxation of water follows a unique decay law in the entire temperature range considered. We found, for all the temperatures, that the time-derivative of the stretched exponential function, $f(t) = -d/dt \exp(-(t/\tau)^\beta)$, allows a very satisfactory fitting of the HD-OKE data: the values of the parameters τ and β obtained at the different temperatures are collected in the box of Fig. 2a.

The comparison of the two fittings in Fig. 3 clearly demonstrates that at 273 K, the difference between a double exponential and a stretched exponential function is hardly detectable owing to the limited time interval explored. At 254 K, on the other hand, the wider time window accessible to the experiment shows that the double exponential function is unable to account for the OKE signal decay in the last 20 ps.

A second important conclusion from the data of Fig. 2a is that the stretching parameter β has a constant value of 0.6 throughout the entire temperature range considered.

As shown in Fig. 4, the temperature dependence of the measured relaxation time τ is well reproduced by a power law of the type $\tau(T) = a(T - T_S)^{-\gamma}$. The best fit is obtained using values of $T_S = 221 \pm 5$ K and $\gamma = 2.2 \pm 0.3$ for the power-law parameters. The T_S we obtain thus coincides with the singularity temperature obtained from measurements of other dynamic and transport properties.

The two important new results obtained from our OKE experiment are the stretched exponential nature of the relaxation and the invariance of the stretching parameter with the temperature. The large stretching effect ($\beta = 0.6$) indicates that water dynamics is characterized by a distribution of different timescales, suggesting the presence of a variety of relaxing structures⁶ whose distribution does not, according to our results, change notably with temperature.

Our data also allow for a comprehensive comparison with the predictions of simple MCT¹⁰ concerning the slow dynamics in liquid water. MCT, which is a dynamic model providing a description of the temporal evolution of the correlation functions related to

density, predicts that the long time decay (α relaxation) follows a stretched exponential law with a temperature-independent stretching parameter. Furthermore, the α relaxation time increases on cooling according to a power law, from which a critical temperature T_S can be derived. These predictions fully agree with our experimental observations. In the simple MCT, T_S represents the temperature of dynamical structural arrest, although recently extended versions propose a new definition of T_S , involving a crossover or avoided-singularity temperature corresponding to a sudden variation of the dynamic nature of the supercooled liquid.

A purely dynamic water model clearly cannot account for the rich and anomalous thermodynamic behaviour of water, which calls for other model types such as the much debated^{16–18} second critical point hypothesis²⁸ (SCP). But the predictions of thermodynamic models usually do not cover dynamic properties, such as the temporal evolution of the correlation functions we have measured. Although some recent computer simulations²⁹ attempt to bridge this gap, we are at present unable to compare our experimental results directly against the predictions of thermodynamic models such as the SCP hypothesis.

On the other hand, our data fully support the MCT scenario, both in terms of the form of the correlation function and in terms of its scaling with temperature. This agreement provides unambiguous evidence that the anomalous behaviour of weakly supercooled water at atmospheric pressure can successfully be described by a fully dynamic model, with no need for a thermodynamic origin for the observed anomalous behaviour. In this picture the singularity temperature T_S , inferred from a range of different dynamic and transport properties of water, is the signature of an avoided dynamical arrest. Despite its strongly hydrogen-bonded nature, weakly supercooled water behaves in this context like a fragile glass-former, as seen for many other molecular liquids. □

Received 31 October 2003; accepted 4 February 2004; doi:10.1038/nature02409.

- Angell, C. A. In *Water: A Comprehensive Treatise* (ed. Franks, F.) Vol. 7 1–81 (Plenum, New York, 1982).
- Angell, C. A. Supercooled water. *Annu. Rev. Phys. Chem.* **34**, 593–630 (1983).
- Mishima, O. & Stanley, H. E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1995).
- Debenedetti, P. G. *Metastable Liquids* Ch. 4 305–335 (Princeton Univ. Press, Princeton, 1996).
- Speedy, R. J. & Angell, C. A. Isothermal compressibility of supercooled water and evidence for thermodynamic singularity at -45°C . *J. Chem. Phys.* **65**, 851–858 (1976).
- Debenedetti, P. G. & Stillinger, F. H. Supercooled liquids and the glass transition. *Nature* **401**, 259–267 (2001).
- Sciortino, F., Gallo, P., Tartaglia, P. & Chen, S.-H. Supercooled water and the kinetic glass transition. *Phys. Rev. E* **54**, 6331–6343 (1996).
- Fabbian, L. *et al.* Molecular mode coupling theory for supercooled liquids: Application to water. *Phys. Rev. E* **60**, 5768–5777 (1999).
- Sciortino, F. Slow dynamics in supercooled water. *Chem. Phys.* **258**, 307–314 (2000).
- Götze, W. & Sjögren, L. Relaxation processes in supercooled liquids. *Rep. Prog. Phys.* **55**, 241–336 (1992).
- Winkler, K., Lindner, J., Bürsing, H. & Vöringer, P. Ultrafast Raman-induced Kerr-effect of water: Single molecule versus collective motions. *J. Chem. Phys.* **113**, 4674–4682 (2000).
- Rønne, C., Åstrand, P.-O. & Keiding, S. R. THz spectroscopy of liquid H_2O and D_2O . *Phys. Rev. Lett.* **82**, 2888–2891 (1999).
- Sokolov, A. P., Hurst, J. & Quitmann, D. Dynamics of supercooled water: Mode-coupling theory approach. *Phys. Rev. B* **51**, 12865–12868 (1995).
- Bellissent-Funel, M. C., Longeville, S., Zanotti, J. M. & Chen, S.-H. Experimental observation of the α relaxation in supercooled water. *Phys. Rev. Lett.* **85**, 3644–3647 (2000).
- Righini, R. Ultrafast optical Kerr effects in liquids and solids. *Science* **262**, 1386–1390 (1993).
- Debenedetti, P. G. & Stanley, H. E. Supercooled and the glassy water. *Phys. Today* **56**, 40–46 (2003).
- Sciortino, F., La Nave, E. & Tartaglia, P. Physics of the liquid-liquid critical point. *Phys. Rev. Lett.* **91**, 155701 (2003).
- Franzese, G., Marqués, M. I. & Stanley, H. E. Intramolecular coupling as a mechanism for a liquid-liquid phase transition. *Phys. Rev. E* **67**, 011103 (2003).
- Castner, E. W., Chang, Y. J., Chu, Y. C. & Walrafen, G. E. The intermolecular dynamics of liquid water. *J. Chem. Phys.* **102**, 653–659 (1995).
- Palese, S., Schilling, L., Miller, R. J. D., Staver, P. R. & Lotshaw, W. T. Femtosecond optical Kerr-effect studies of water. *J. Phys. Chem.* **98**, 6308–6316 (1994).
- Palese, S., Mukamel, S., Miller, R. J. D. & Lotshaw, W. T. Interrogation of vibrational structure and line broadening of liquid water by Raman-induced Kerr-effect measurements within the multimode Brownian oscillator model. *J. Phys. Chem.* **100**, 10380–10388 (1996).
- Foggi, P., Bellini, M., Kien, D. P., Vercuque, I. & Righini, R. Relaxation dynamics of water and HCl aqueous solutions measured by time-resolved optical Kerr effect. *J. Phys. Chem.* **101**, 7029–7035 (1997).

23. Winkler, K., Lindner, J. & Vöhringer, P. Low frequency depolarized Raman-spectral density of liquid water from femtosecond optical Kerr-effect measurements: Lineshape analysis of restricted translational modes. *Phys. Chem. Chem. Phys.* **4**, 2144–2155 (2002).
24. Bartolini, P., Ricci, M., Torre, R. & Righini, R. Diffusive and oscillatory dynamics of iodobenzene measured by femtosecond optical Kerr effect. *J. Chem. Phys.* **110**, 8653–8662 (1999).
25. Ricci, M. A., Ruocco, G. & Sampoli, M. Raman spectra of water in the translational and librational regions. III — Temperature evolution by computer simulation with the TIP4P potential. *Mol. Phys.* **67**, 19–31 (1989).
26. Torre, R., Bartolini, P. & Pick, R. M. Time-resolved optical Kerr effect in a fragile glass-forming liquid, salol. *Phys. Rev. E* **57**, 1912–1920 (1998).
27. Torre, R., Bartolini, P., Ricci, M. & Pick, R. M. Time-resolved optical Kerr effect in a fragile glass-forming liquid: Test of different mode coupling theory aspects. *Europhys. Lett.* **52**, 324–329 (2000).
28. Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. Phase behaviour of metastable water. *Nature* **360**, 324–328 (1992).
29. Sastry, S. & Angell, C. A. Liquid-liquid phase transition in supercooled silicon. *Nature Mater.* **2**, 739–743 (2003).

Acknowledgements This work was supported by INFN, by the FIRB and COFIN programmes of Italian MIUR, and by an EC grant. We thank F. Sciortino, C. A. Angell and G. Ruocco for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.T. (torre@lens.unifi.it).

High-mobility ultrathin semiconducting films prepared by spin coating

David B. Mitzl, Laura L. Kosbar, Conal E. Murray, Matthew Copel & Ali Afzali

IBM T. J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA

The ability to deposit and tailor reliable semiconducting films (with a particular recent emphasis on ultrathin systems) is indispensable for contemporary solid-state electronics^{1–3}. The search for thin-film semiconductors that provide simultaneously high carrier mobility and convenient solution-based deposition is also an important research direction, with the resulting expectations of new technologies (such as flexible or wearable computers, large-area high-resolution displays and electronic paper) and lower-cost device fabrication^{4–11}. Here we demonstrate a technique for spin coating ultrathin (~50 Å), crystalline and continuous metal chalcogenide films, based on the low-temperature decomposition of highly soluble hydrazinium precursors. We fabricate thin-film field-effect transistors (TFTs) based on semiconducting $\text{SnS}_{2-x}\text{Se}_x$ films, which exhibit n-type transport, large current densities ($>10^5 \text{ A cm}^{-2}$) and mobilities greater than $10 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ —an order of magnitude higher than previously reported values for spin-coated semiconductors. The spin-coating technique is expected to be applicable to a range of metal chalcogenides, particularly those based on main group metals, as well as for the fabrication of a variety of thin-film-based devices (for example, solar cells¹², thermoelectrics¹³ and memory devices¹⁴).

Recent research on solution-processed semiconducting films has primarily focused on organic systems^{4–9}, with mobilities as high as $0.89 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ reported for p-type TFT channels composed of solution-processed pentacene⁶. Analogous n-type materials have proved more challenging, and have generally yielded much lower mobilities^{8,9}. Although promising with regard to processing, cost and weight considerations, the weak van der Waals interaction between organic moieties imposes an upper

bound on mobility⁹, thereby limiting organic semiconductors to lower-end applications. Spin-coated organic–inorganic hybrid films have also recently been used as the semiconducting element in TFTs^{10,11}, yielding saturation regime mobilities as high as $1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (ref. 11). Although promising, current tin(II)-iodide-based hybrids are air-sensitive, requiring all processing and characterization to be performed under rigorously inert atmospheric conditions. Additionally, although the pentacene and tin(II)-iodide-based systems are p-type, n-type analogues are also desirable to enable applications involving complementary logic.

Metal chalcogenide films provide an opportunity for higher mobility than the organic and hybrid semiconductors, as a result

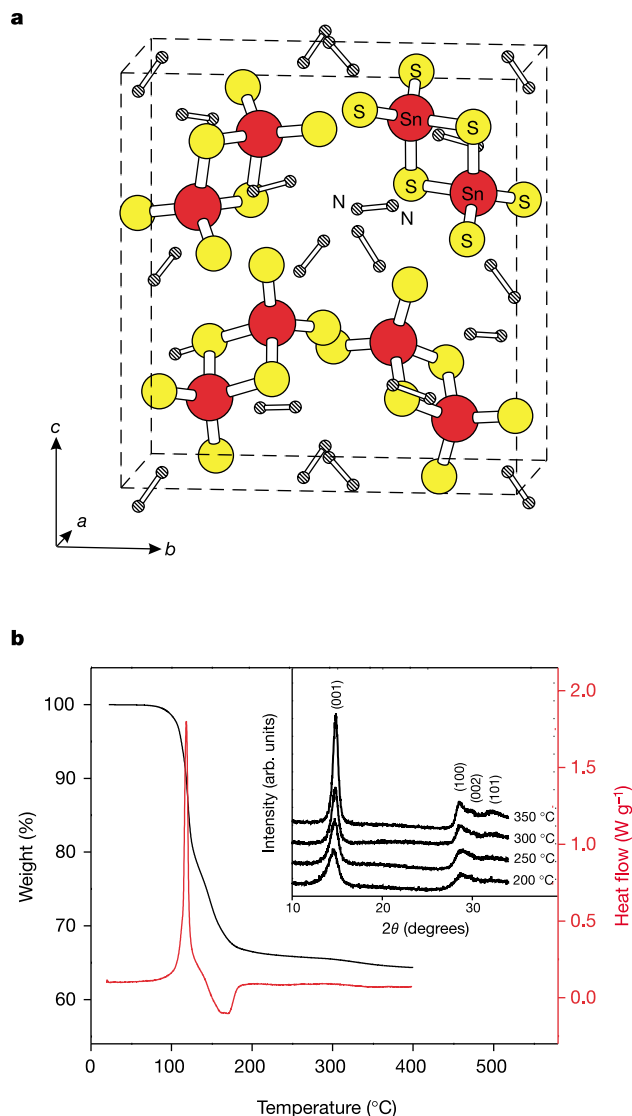


Figure 1 Crystal structure and thermal properties of the hydrazinium precursor. **a**, Crystal structure of $(\text{N}_2\text{H}_5)_4\text{Sn}_2\text{S}_6$, determined using single-crystal X-ray diffraction. Hydrogen atoms are not shown for clarity. Unit cell (outlined by dashed lines): orthorhombic ($P2_12_12_1$), $a = 8.5220(5) \text{ \AA}$, $b = 13.6991(8) \text{ \AA}$, $c = 14.4102(9) \text{ \AA}$, $V = 1,682.3(2) \text{ \AA}^3$ and $Z = 4$. Further details of the structure determination for both $(\text{N}_2\text{H}_5)_4\text{Sn}_2\text{S}_6$ and the analogous selenide precursor will be published elsewhere. **b**, Thermogravimetric analysis (TGA) and differential scanning calorimetry (DSC) scans for the $(\text{N}_2\text{H}_5)_4\text{Sn}_2\text{S}_6$ precursor (2°C min^{-1} ramp rate). Exothermic peaks are down along the heat flow axis. Inset, the powder X-ray diffraction pattern for the precursor after a TGA run terminated at various temperatures (200–350 $^\circ\text{C}$) (listed along the right side of the plot), demonstrating good agreement with the expected diffraction pattern for SnS_2 (berndite, PDF 23-0677; the reflection indices are shown in parentheses).

of the more covalent nature of the chalcogenide framework, as well as for identifying n-type semiconductors. Reported ambient temperature bulk mobilities of metal chalcogenides include SnSe_2 ($27 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$; n-type)¹⁵, SnS_2 ($18 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$; n-type)¹⁶, CdS ($250 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$; n-type), CdSe ($650 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$; n-type), ZnSe ($600 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$; n-type), and ZnTe ($100 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$; p-type)¹⁷. Unfortunately, although the potential for higher mobility exists, the increased covalency of the extended metal chalcogenide systems reduces their solubility, rendering simple and low-cost film deposition a significant challenge (metal chalcogenide films are generally deposited using thermal evaporation, chemical vapour deposition or sputtering techniques). A few solution-based deposition processes, such as galvanic deposition¹⁸ and chemical bath deposition¹⁹, have successfully been used to form films for device applications. The chemical bath technique, in particular, has yielded cadmium selenide films (after 40 min of multiple dipping and a final anneal at 400°C) with mobility values as high as $15 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (ref. 19). However, although these techniques are solution-based, they are generally not high-throughput processes like spin coating, printing and stamping, which require a truly soluble precursor for the chalcogenide and a suitable solvent. To this end, CdSe films have recently been cast²⁰ with a mobility as high as $1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (similar to the organic and hybrid films), using a soluble metal chalcogenide

precursor consisting of organic derivatized CdSe nanocrystals.

The current study examines the use of hydrazine as a solvent for main group metal chalcogenides (for example, $\text{SnS}_{2-x}\text{Se}_x$) and the subsequent processing and characterization of crystalline thin films based on these solutions. The solvent properties of hydrazine with a wide variety of inorganic substances have been previously examined²¹, with negligible solubility reported for many of the binary and ternary chalcogenides. The simple approach described here uses excess chalcogen (S or Se) to directly improve the solubility and film-forming properties of selected main group metal chalcogenides in hydrazine, as a result of the formation of highly soluble hydrazinium-based precursors that can be solution processed into thin-film form and decomposed to the starting metal chalcogenide at low temperature. Continuous crystalline semiconducting films as thin as $50 \text{ \AA}$ can be formed by spin coating using this process, enabling the formation of high-performance spin-coated channel layers for

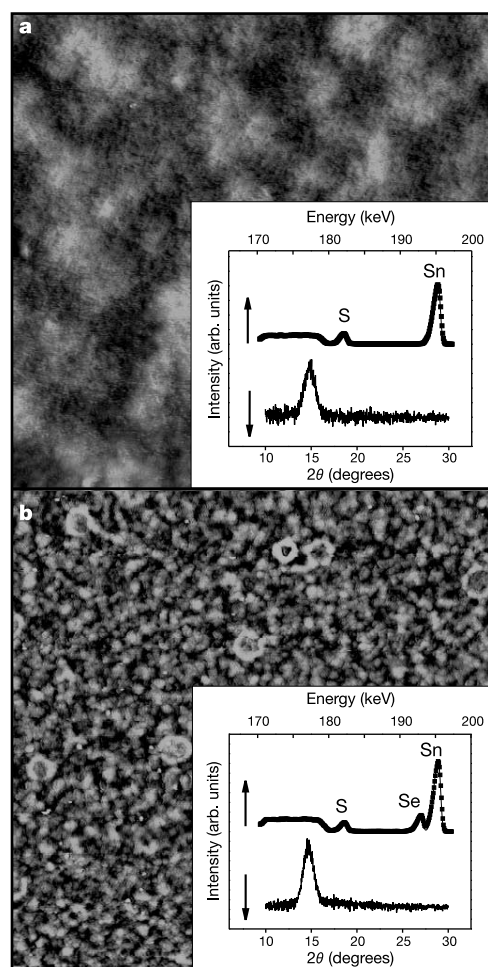


Figure 2 Chemical and structural properties of solution-processed chalcogenide films. **a, b**, Atomic force microscope (AFM) images of the chalcogenide films after annealing, spun using solutions (1) and (2), respectively (see Methods). Each panel represents a $2 \mu\text{m} \times 2 \mu\text{m}$ scan. Inset, powder X-ray diffraction pattern (bottom trace in each inset) and MEIS spectrum (top trace) for each film. Ratios of integrated MEIS peak intensities establish film composition, while peak width reflects the film thickness²⁵.

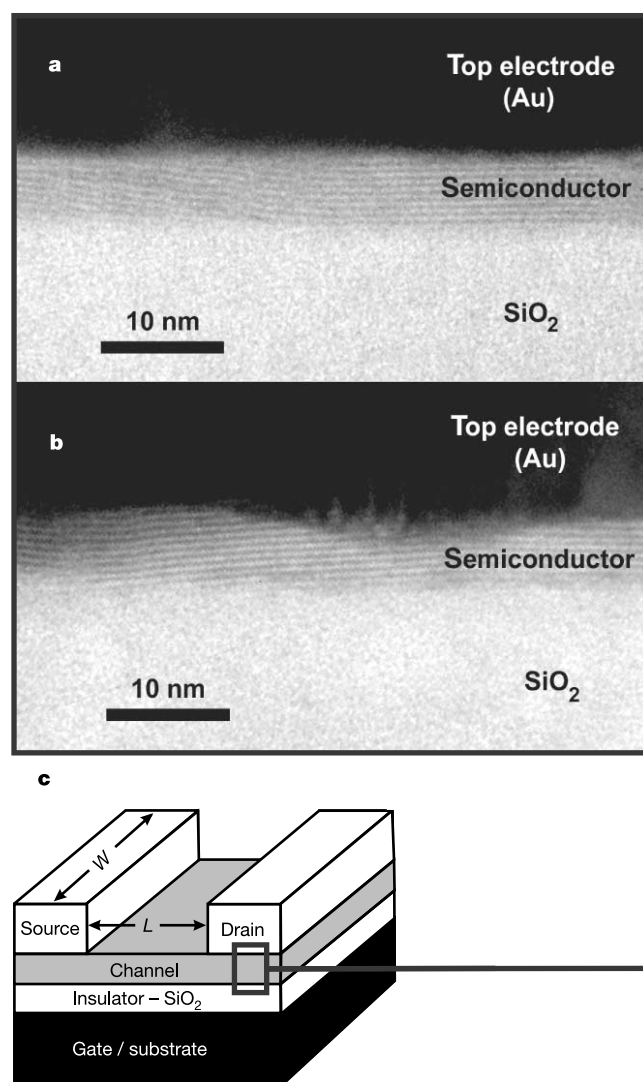
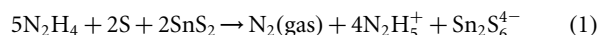


Figure 3 Diagram of transistor, with TEM cross-section of spin-coated semiconductors. **a, b**, Cross-sectional TEM images of devices based on solutions (1) and (2), respectively. **c**, The thin-film field-effect transistor (TFT) device (not shown to scale) consists of a heavily n-doped silicon wafer as the substrate/gate, a $2.1 \text{ k\AA}$ thermally grown SiO_2 gate dielectric, the spin-coated chalcogenide channel layer (deposited as described in the text), and patterned electron-beam-evaporated Au source and drain electrodes. The channel length (L), defined as the distance between source and drain electrodes, ranged from 5 to $95 \mu\text{m}$, whereas the channel width (W) ranged from 250 to $1,500 \mu\text{m}$ in the devices examined.

TFTs with all processing performed at or below 300 °C.

The soluble SnS_2 hydrazinium precursor is isolated by stirring tin(IV) sulphide and sulphur in hydrazine under ambient conditions (see Methods). Dissolution primarily occurs through the overall reaction:



Evaporation of the solution yields $(\text{N}_2\text{H}_5)_4\text{Sn}_2\text{S}_6$ crystals, which consist of anionic dimers of edge-sharing SnS_4 tetrahedra ($\text{Sn}_2\text{S}_6^{4-}$) alternating with hydrazinium cations (Fig. 1a). The observed crystal structure is similar to that found²² in $\text{K}_4\text{Sn}_2\text{Se}_6$, with the replacement of the alkali metal cation by a small volatile hydrazinium species that can be decomposed from the material using a low-temperature anneal. As seen in Fig. 1b, $(\text{N}_2\text{H}_5)_4\text{Sn}_2\text{S}_6$ decomposes endothermally (237 kJ mol^{-1}) with an onset at $\sim 114^\circ\text{C}$, followed by a second broad exothermic process at slightly higher temperature, ultimately leading to condensation of the $\text{Sn}_2\text{S}_6^{4-}$ dimers into the extended SnS_2 framework²³. The observed weight loss during the decomposition process (34.6% at 300°C) agrees with the expected value (34.9%) for the formation of SnS_2 . However, the weight loss curve is not completely flat above the decomposition transition, suggesting that sulphur loss and deviation from ideal SnS_2 stoichiometry may occur at increased temperature. The weight loss process is essentially completed, and crystallinity begins to appear, by $\sim 200^\circ\text{C}$ (Fig. 1b inset), highlighting the low-temperature nature of this process. An analogous selenide precursor, $(\text{N}_2\text{H}_4)_3(\text{N}_2\text{H}_5)_4\text{Sn}_2\text{Se}_6$, has also been isolated from a tin(IV) selenide, selenium and hydrazine solution. The structure of this precursor is similar to the sulphide analogue, but with the incorporation of additional neutral hydrazine molecules among the $\text{Sn}_2\text{Se}_6^{4-}$ dimers.

Chalcogenide films were spin coated in air onto thermally oxidized silicon substrates (see Methods), from one of two solutions prepared from freshly distilled hydrazine: (1) 20 mg SnS_2 and 4 mg S

in 1.6 ml hydrazine; or (2) 14 mg SnS_2 , 8 mg SnSe_2 and 6 mg S in 1.9 ml hydrazine. Final annealing/decomposition treatments were in a nitrogen atmosphere at either 295°C (solution (1)) or 270°C (solution (2)). Note that the relatively low temperatures used during fabrication are already compatible with selected flexible plastic substrates (for example, Kapton), whereas lower-temperature plastics might also be compatible if pulsed-laser annealing were used for the final heating step, as has been previously demonstrated for the preparation of polysilicon films²⁴. The chemical compositions (and average film thicknesses) of films based on solutions (1) and (2), as determined by medium-energy ion scattering (MEIS)²⁵, are $\text{SnS}_{1.8(1)}$ ($50(5) \text{ \AA}$) and $\text{SnS}_{1.4(1)}\text{Se}_{0.5(1)}$ ($42(5) \text{ \AA}$), respectively (Fig. 2, insets). The MEIS results suggest a slight chalcogen deficiency in the films, relative to the ideal 2:1 S/Se:Sn ratio. The two thin films are continuous and relatively uniform, as seen in atomic force microscope (AFM) images (Fig. 2). Inclusion of tin(IV) selenide in the starting solution appears to significantly affect the film surface morphology, leading to r.m.s. roughnesses of $14 \text{ \AA}$ (Fig. 2b) versus $6 \text{ \AA}$ (Fig. 2a). The single relatively broad X-ray diffraction peak from each of the ultrathin films (Fig. 2, insets) matches the (001) reflection of bulk-prepared SnS_2 (that is, Fig. 1b, inset) and verifies the film crystallinity. Note that the small size of the hydrazinium cation (and the correspondingly small film volume loss during decomposition) probably contributes to the high quality of the resulting $\text{SnS}_{2-x}\text{Se}_x$ films and the ability to achieve ultrathin dimensions.

TFTs based on the spin-coated chalcogenide films have been fabricated with the architecture shown in Fig. 3. In contrast to the organic–inorganic hybrid semiconductors, which must be maintained in a rigorously inert atmosphere, much of the spin-coated chalcogenide device fabrication is performed in air—with the exception of the final channel film anneal (performed in nitrogen) and the gold source/drain contact deposition (performed in vacuum). Transmission electron microscope (TEM) images of a

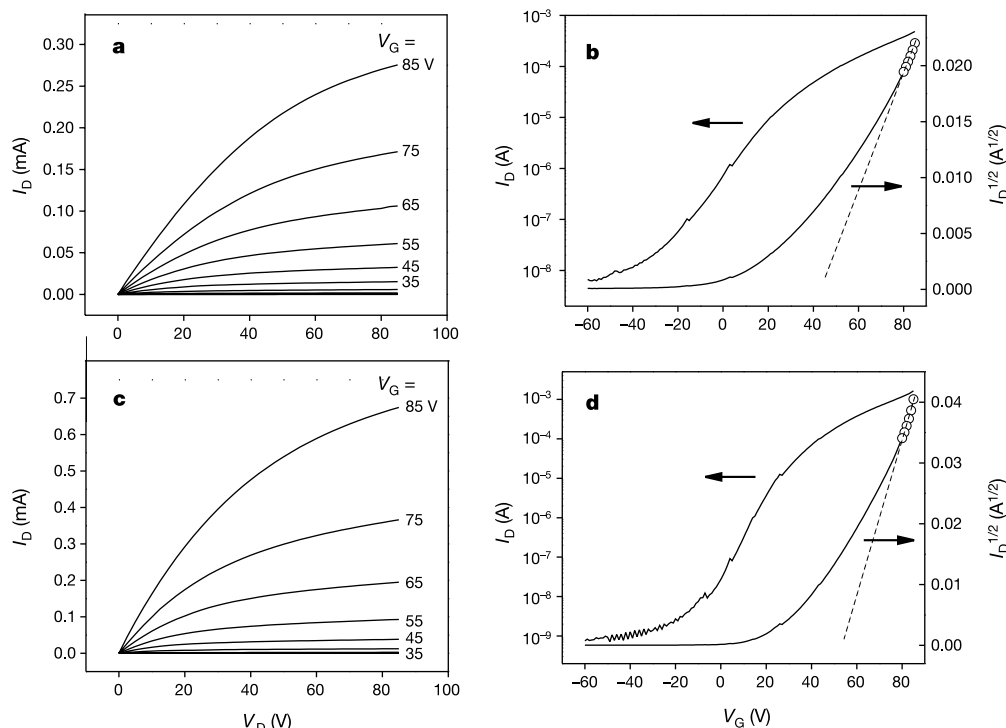


Figure 4 Device characteristics for spin-coated chalcogenide channel layers. **a**, Drain current, I_D , versus drain voltage, V_D , as a function of gate voltage, V_G , for a spin-coated tin sulphide channel layer, deposited from solution (1), and having a channel length $L = 14 \mu\text{m}$ and channel width $W = 500 \mu\text{m}$. **b**, Plots of I_D and $I_D^{1/2}$ versus V_G at constant

$V_D = 85 \text{ V}$, used to calculate current modulation, $I_{\text{on}}/I_{\text{off}}$, and saturation regime mobility, μ_{sat} , for the tin sulphide channel. **c**, Plots of I_D versus V_D , as a function of V_G , for a spin-coated $\text{SnS}_{1.4}\text{Se}_{0.5}$ channel with $L = 14 \mu\text{m}$ and $W = 250 \mu\text{m}$, deposited from solution (2). **d**, Plots of I_D and $I_D^{1/2}$ versus V_G at constant $V_D = 85 \text{ V}$ for the $\text{SnS}_{1.4}\text{Se}_{0.5}$ film.

device cross-section (Fig. 3a and b) demonstrate the ultrathin, crystalline nature of the films, with average film thicknesses of ~ 50 Å (consistent with the MEIS results)—that is, only ~ 8 – 10 unit cells thick. Substantial preferred orientation is visible within the films, with the $\text{SnS}_{2-x}\text{Se}_x$ layers oriented nominally parallel to the substrate surface (corresponding to the fringes observed in the TEM images). The average interlayer (fringe) separation is $6(1)$ Å, consistent with the (001) layer separation of the $\text{SnS}_{2-x}\text{Se}_x$ structures²³.

A representative plot of drain current, I_D , versus drain voltage, V_D , is shown in Fig. 4a as a function of the applied gate voltage, V_G , for a TFT with a tin sulphide channel formed from solution (1). The device operates as an n-channel transistor, operating in accumulation mode upon application of a positive gate bias. Application of a negative gate bias depletes the channel of electrons and shuts the device off. At low V_D , the TFT shows typical transistor behaviour as I_D increases linearly with V_D . Current saturation, with a small ohmic component, is observed at high V_D as the accumulation layer is pinched off near the drain electrode. Current modulation ($I_{\text{ON}}/I_{\text{OFF}}$) and saturation regime field-effect mobility (μ_{sat}) are calculated from the plot of I_D and $I_D^{1/2}$ versus V_G (Fig. 4b)⁹, yielding $I_{\text{ON}}/I_{\text{OFF}} = 8 \times 10^4$ and $\mu_{\text{sat}} = 0.89 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, respectively, for a -60 to 85 V gate sweep and $V_D = 85$ V. Note that use of a thinner or higher dielectric constant gate insulator (relative to the 2.1 -kÅ SiO_2 layer currently used) is expected to enable significant reduction of the device operating voltage²⁶. The linear regime mobility derived from the plots in Fig. 4a, $\mu_{\text{lin}} = 0.32 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, is lower than the saturation regime value, as also typically observed for organic and organic–inorganic hybrid semiconductors^{9,27}.

Introduction of SnSe_2 into solution (2) dramatically improves the characteristics of devices formed (Figs 4c and d), yielding $I_{\text{ON}}/I_{\text{OFF}} > 10^6$, with a large peak current density of $\sim 10^5 \text{ A cm}^{-2}$, and mobilities of $\mu_{\text{sat}} = 12.0 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $\mu_{\text{lin}} = 2.4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. These mobility values are representative of the devices measured, and are approximately an order of magnitude higher than previous results for spin-coated films. In a typical run involving two substrates and a total of 17 devices measured (biased to 85 V), μ_{sat} ranged from 6.4 to $13.3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (only three devices, with the largest channel lengths examined, had a mobility lower than $9 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$). Devices biased up to 85 V for 100 cycles (in a nitrogen atmosphere) exhibit minimal, if any, mobility or on–off ratio degradation. The improved device characteristics observed for the selenide-containing devices probably arise from the differences in grain structure and electronic structure induced in the films by the sulphur–selenium substitution.

Note that TFTs based on the flow-directed assembly (not a high-throughput spin-coating process) of Si nanowires and CdS nanoribbons have also recently been reported to yield high mobility²⁸. But these results are for collections of discrete nanowires across the electrodes, rather than for a continuous film, and the quoted mobilities of $>100 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ are based on an effective channel width calculated using the nanowire diameter multiplied by the number of wires crossing the channel. Use of the true channel width of the device (the most useful comparison for thin-film devices) would, from the reported effective ($\sim 6.5 \text{ m}^{-1}$) and uncorrected ($\sim 0.09 \text{ S m}^{-1}$) transconductances for the Si nanowire devices, yield a mobility approximately two orders of magnitude smaller than the reported values and lower than those presented here.

Although our spin-coating technique is described here for $\text{SnS}_{2-x}\text{Se}_x$, selected other metal chalcogenides (particularly main group metal systems) with useful metallic, semiconducting or insulating properties are also likely to be suitable for deposition using the hydrazinium-precursor process. For example, preliminary results suggest that Sb_2X_3 ($\text{X} = \text{S}, \text{Se}$ or Te) may be deposited using an analogous process. Higher mobilities can therefore be expected as other metal chalcogenide systems are explored. The process can also be generalized to include mixtures of hydrazine with other less toxic solvents (for example, we recently demonstrated a 50:50

water:hydrazine mixture for spin coating chalcogenide films: D.B.M., unpublished results). Finally, the described spin-coated films may be useful in the preparation of a wide range of thin-film and multilayer devices, including solar cells¹² and thermoelectric¹³ and memory devices¹⁴, as well as devices taking advantage of the ultrathin films for quantum confinement effects. □

Methods

Synthesis of $(\text{N}_2\text{H}_5)_4\text{Sn}_2\text{S}_6$

Pale yellow crystals of $(\text{N}_2\text{H}_5)_4\text{Sn}_2\text{S}_6$ are formed by dissolving (in a nitrogen atmosphere, over several hours, with stirring) 0.183 g of SnS_2 (1 mmol) in 2 ml of hydrazine and 0.064 g of S (2 mmol), ultimately yielding a yellow solution. Vigorous bubbling during dissolution supports the gas generation indicated in equation (1). The highly toxic hydrazine was handled using appropriate protective equipment to prevent physical contact with either vapour or liquid. Evaporating the solution to dryness under flowing nitrogen gas over a period of several hours, as well as under vacuum in the dry-box antechamber, leads to the formation of approximately 0.274 g (0.49 mmol) of the crystalline yellow product, $(\text{N}_2\text{H}_5)_4\text{Sn}_2\text{S}_6$. Chemical analysis: observed, 19.8% N, 3.5% H; calculated, 19.9% N, 3.6% H.

Thin film deposition

Chalcogenide films are spin coated onto thermally oxidized silicon substrates, cleaned using a 'piranha' process (hydrogen peroxide:sulphuric acid in 1:4 volume ratio), from one of two solutions prepared from freshly distilled hydrazine: (1) 20 mg SnS_2 and 4 mg S in 1.6 ml hydrazine; or (2) 14 mg SnS_2 , 8 mg SnSe_2 and 6 mg S in 1.9 ml hydrazine. The solid reactants completely dissolved after stirring at room temperature in a nitrogen-filled vial (after 6 h and 2 h for solutions (1) and (2), respectively). Spin coating was performed in air by flooding the substrate surface with the clear yellow solution (passed through a 0.2 - μm filter) and spinning at $3,500$ r.p.m. for 45 s. The films were then immediately dried in air for 5 min at 120°C on a preheated hot plate, followed by a final 20 min anneal at either 295°C (solution (1)) or 270°C (solution (2)) in a nitrogen-filled dry box with oxygen and water levels maintained below 1 p.p.m. Annealing temperatures were verified on the top surface of a silicon substrate positioned on the hot plate with an attached K-type (chromel–alumel) thermocouple.

Characterization

Thermogravimetric analysis (TGA) and differential scanning calorimetry (DSC) scans were performed using TA Instruments TGA-2950 and DSC-2010 systems, respectively. TGA was run in a flowing nitrogen atmosphere. For the DSC scan, non-hermetically sealed aluminium capsules were used in a flowing helium atmosphere. The AFM images were collected using a Digital Instruments Dimension 3100 scanning probe microscope, operated in tapping mode with an etched silicon cantilever. MEIS analysis was performed with 200 -keV protons using an electrostatic energy analyser at a scattering angle of 120° . The large scattering angle was chosen to minimize overlap between the Sn and Se peaks. The system was calibrated to better than 5% accuracy using a SiO_2 film of known thickness. TEM images were taken using a JEOL 4000FX TEM operated at 400 kV. TFT devices were tested using a Hewlett-Packard 4145B semiconductor analyser in a nitrogen-filled dry box. Whereas the devices appeared to be stable in air (at least for several days) without being electrically driven, operation in air severely shortened the device lifetime.

Received 3 November 2003; accepted 3 February 2004; doi:10.1038/nature02389.

- Zaslavsky, A. *et al.* Ultrathin silicon-on-insulator vertical tunneling transistor. *Appl. Phys. Lett.* **83**, 1653–1655 (2003).
- Trivedi, V. P. & Fossum, J. G. Scaling fully depleted SOI CMOS. *IEEE Trans. Electron Devices* **50**, 2095–2103 (2003).
- Forrest, S. R. Ultrathin organic films grown by organic molecular beam deposition and related techniques. *Chem. Rev.* **97**, 1793–1896 (1997).
- Huitema, H. E. A. *et al.* Active-matrix displays driven by solution processed polymeric transistors. *Adv. Mater.* **14**, 1201–1204 (2002).
- Mushrush, M., Facchetti, A., Lefenfeld, M., Katz, H. E. & Marks, T. J. Easily processable phenylene-thiophene-based organic field-effect transistors and solution-fabricated nonvolatile transistor memory elements. *J. Am. Chem. Soc.* **125**, 9414–9423 (2003).
- Afzali, A., Dimitrakopoulos, C. D. & Breen, T. L. High-performance, solution-processed organic thin film transistors from a novel pentacene precursor. *J. Am. Chem. Soc.* **124**, 8812–8813 (2002).
- Sirringhaus, H. *et al.* High-resolution inkjet printing of all-polymer transistor circuits. *Science* **290**, 2123–2126 (2000).
- Katz, H. E. *et al.* A soluble and air-stable organic semiconductor with high electron mobility. *Nature* **404**, 478–481 (2000).
- Dimitrakopoulos, C. D. & Malenfant, P. R. L. Organic thin film transistors for large area electronics. *Adv. Mater.* **14**, 99–117 (2002).
- Kagan, C. R., Mitzi, D. B. & Dimitrakopoulos, C. D. Organic–inorganic hybrid materials as semiconducting channels in thin-film field-effect transistors. *Science* **286**, 945–947 (1999).
- Mitzi, D. B. & Kagan, C. R. in *Thin-Film Transistors* (eds Kagan, C. R. & Andry, P.) 475–513 (Marcel Dekker, New York, 2003).
- Britt, J. & Ferekides, C. Thin-film CdS/CdTe solar cell with 15.8% efficiency. *Appl. Phys. Lett.* **62**, 2851–2852 (1993).
- Kim, Y. *et al.* Structural and thermoelectric transport properties of Sb_2Te_3 thin films grown by molecular beam epitaxy. *J. Appl. Phys.* **91**, 715–718 (2002).
- Kyratsi, T., Chrissafis, K., Wachter, J., Paraskevopoulos, K. M. & Kanatzidis, M. G. KSb_2S_6 : A wide bandgap phase-change material for ultra high density rewritable information storage. *Adv. Mater.* **15**, 1428–1431 (2003).

15. Domingo, G., Itoga, R. S. & Kannewurf, C. R. Fundamental optical absorption in SnS_2 and SnSe_2 . *Phys. Rev.* **143**, 536–541 (1966).
16. Shibata, T., Muranushi, Y., Miura, T. & Kishi, T. Electrical characterization of 2H- SnS_2 single crystals synthesized by the low temperature chemical vapor transport method. *J. Phys. Chem. Solids* **52**, 551–553 (1991).
17. Streetman, B. G. *Solid State Electronic Devices* 443 (Prentice-Hall, New Jersey, 1980).
18. McCandless, B. E., Mondal, A. & Birkmire, R. W. Galvanic deposition of cadmium sulfide thin films. *Sol. Energy Mater. Sol. Cells* **36**, 369–379 (1995).
19. Gan, F. Y. & Shih, I. Preparation of thin-film transistors with chemical bath deposited CdSe and CdS thin films. *IEEE Trans. Electron Devices* **49**, 15–18 (2002).
20. Ridley, B. A., Nivi, B. & Jacobson, J. M. All-inorganic field effect transistors fabricated by printing. *Science* **286**, 746–749 (1999).
21. Welsh, T. W. B. & Broderson, H. J. Anhydrous hydrazine. III. Anhydrous hydrazine as a solvent. *J. Am. Chem. Soc.* **37**, 816–824 (1915).
22. Eisenmann, B. & Hansa, J. Crystal structure of tetrapotassium hexaselenodistannate, $\text{K}_4\text{Sn}_2\text{Se}_6$. *Z. Kristallogr.* **203**, 299–300 (1993).
23. Palosz, B., Steurer, W. & Schulz, H. Refinement of SnS_2 polytypes 2H, 4H and 18R. *Acta Crystallogr. B* **46**, 449–455 (1990).
24. Carey, P. G., Smith, P. M., Theiss, S. D. & Wickboldt, P. Polysilicon thin film transistors fabricated on low temperature plastic substrates. *J. Vac. Sci. Technol. A* **17**, 1946–1949 (1999).
25. Van der Veen, J. F. Ion beam crystallography of surfaces and interfaces. *Surf. Sci. Rep.* **5**, 199–287 (1985).
26. Dimitrakopoulos, C. D., Purushothaman, S., Kyriassis, J., Callegari, A. & Shaw, J. M. Low-voltage organic transistors on plastic comprising high-dielectric constant gate insulators. *Science* **283**, 822–824 (1999).
27. Mitzi, D. B., Dimitrakopoulos, C. D. & Kosbar, L. L. Structurally tailored organic-inorganic perovskites: Optical properties and solution-processed channel materials for thin-film transistors. *Chem. Mater.* **13**, 3728–3740 (2001).
28. Duan, X. *et al.* High-performance thin-film transistors using semiconductor nanowires and nanoribbons. *Nature* **425**, 274–278 (2003).

Acknowledgements We thank C. Dimitrakopoulos for discussions, and D. Martinez, S. J. Chey, D. DiMilia, K. Fogel and T. Graham for technical assistance with the preparation of the substrates.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.B.M. (dmitzi@us.ibm.com).

A 'snowball Earth' climate triggered by continental break-up through changes in runoff

Yannick Donnadieu¹, Yves Godd  ris², Gilles Ramstein¹, Anne N  d  lec² & Joseph Meert³

¹Laboratoire des Sciences du Climat et de l'Environnement, CNRS-CEA, 91191, Gif sur Yvette, France

²Laboratoire des M  canismes et Transferts en G  ologie, UMR 5563, CNRS – Universit   Paul Sabatier – IRD, 31400, Toulouse, France

³Department of Geological Sciences, University of Florida, Gainesville, Florida 32611, USA

Geological and palaeomagnetic studies indicate that ice sheets may have reached the Equator at the end of the Proterozoic eon, 800 to 550 million years ago^{1,2}, leading to the suggestion of a fully ice-covered 'snowball Earth'^{3,4}. Climate model simulations indicate that such a snowball state for the Earth depends on anomalously low atmospheric carbon dioxide concentrations^{5,6}, in addition to the Sun being 6 per cent fainter than it is today. However, the mechanisms producing such low carbon dioxide concentrations remain controversial^{7,8}. Here we assess the effect of the palaeogeographic changes preceding the Sturtian glacial period, 750 million years ago, on the long-term evolution of atmospheric carbon dioxide levels using the coupled climate⁹–geochemical¹⁰ model GEOCLIM. In our simulation, the continental break-up of Rodinia leads to an increase in runoff and hence consumption of carbon dioxide through continental weathering that decreases atmospheric carbon dioxide concen-

trations by 1,320 p.p.m. This indicates that tectonic changes could have triggered a progressive transition from a 'greenhouse' to an 'icehouse' climate during the Neoproterozoic era. When we combine these results with the concomitant weathering effect of the voluminous basaltic traps erupted throughout the break-up of Rodinia¹¹, our simulation results in a snowball glaciation.

Long-term ($\geq 10^6$ yr) evolution of the partial pressure of atmospheric CO_2 (p_{CO_2}) is controlled by the relative importance of degassing through volcanic and mid-ocean-ridge processes and the consumption of CO_2 through continental silicate weathering¹². Any long-term decrease in atmospheric CO_2 can be induced either by a decrease in solid Earth degassing rate, or by an increase in the weathering of continental surfaces. Little is known about the evolution of the degassing rate over the Neoproterozoic era, so linking the global Neoproterozoic cold climate to low degassing rate would be extremely speculative. On the other hand, the sink of CO_2 via continental silicate weathering depends on a variety of parameters, such as the air temperature, continental runoff¹³, vegetation¹⁴, and mechanical weathering¹⁵. The long-term evolution of some of these parameters can be evaluated within the particular context of the Neoproterozoic. The tectonic environment is indeed characterized by the dispersal of continental plates through the break-up of the Rodinia supercontinent between 800 and 700 Myr ago. This break-up may have had two major effects on the sink of CO_2 through continental silicate weathering. First, the break-up of Rodinia is heralded by, and accompanied by, the eruption of large basaltic provinces¹¹, resulting in an increase of the weatherability of the continental surface and consumption of atmospheric CO_2 on the 10^6 -yr timescale^{8,16}. More importantly, the break-up of a supercontinent into several smaller plates will result in an increase of precipitation and runoff over the continental masses, owing to an increase in the sources of moisture along continental borders. This process can boost continental silicate weathering and consume atmospheric CO_2 . Hence, precise quantitative evaluation of changes in atmospheric p_{CO_2} due to palaeogeographic changes requires a sophisticated approach in which the weathering rates are spatially resolved.

We have interfaced a long-term global carbon cycle model to a coupled ocean–atmosphere model (GEOCLIM: 'geological time-scales climate'). This approach allows us to account for the spatial variations (for example, longitude and latitude) of the climatic parameters used in the geochemical model to estimate continental weathering rates: runoff and temperature. We show here that the dispersal of the supercontinent results in an enhanced consumption of atmospheric CO_2 , of the order of 1,320 p.p.m., and in a significant cooling of global temperatures ($\sim 8^\circ\text{C}$). We further demonstrate that experiments accounting for the palaeogeographic effect on weathering rates together with the weathering of the large magmatic provinces produce conditions able to trigger a full snowball glaciation at 750 Myr ago.

The climate module of our coupled geochemical–climate model enabled us to address the following requirements. First, the model must explicitly simulate the hydrologic cycle, and second, the model must have a fast turnaround time. The coupled climate model CLIMBER-2 meets these requirements since, first, water fluxes are explicitly resolved, and second, the design of CLIMBER-2 results in a low computational cost, enabling very long simulations (a few thousand years). Briefly, the atmospheric module is a 2.5-dimensional statistical-dynamical model and has a resolution of 10° in latitude and approximately 51° in longitude. The ocean module describes the zonally averaged characteristics for the ocean realm with a latitudinal resolution of 2.5° . The CLIMBER-2 model is fully described in ref. 9, and successfully simulates the last glacial/interglacial cycle¹⁷; it has been used previously to investigate Neoproterozoic climates¹⁸. In order to determine the atmospheric CO_2 evolution, we use the geochemical COMBINE model¹⁰, which is a box-model including the mathematical description of the global

biogeochemical cycles of carbon, phosphorus, alkalinity and oxygen (see Methods).

Figure 1a, b displays the two palaeogeographies used in this study; ~800 Myr ago (denoted SC, supercontinent configuration) and ~750 Myr ago (denoted DC, dispersed configuration), illustrating the dispersal of the supercontinent^{19,20}. The numerical experiments are described in the Methods section. Changes in the continental distribution, from SC to DC, trigger a large decrease in atmospheric CO₂ levels from 1,830 p.p.m. to 510 p.p.m. for the standard runs with the present-day degassing rate (Fig. 2). The climate of the SC experiment at 1,830 p.p.m. is rather cold, given the reduction of 6% of the incoming solar radiation. Indeed, the mean global temperature is 10.8 °C (against 15 °C for the modern one). Nevertheless, regions of the Earth where temperatures are below the freezing point are mainly located over the polar oceans, and extend down to 60° latitude in both hemispheres (not shown). In contrast, in the DC climate simulation at 510 p.p.m., the freezing temperatures migrate towards mid-latitudes (40–45°; not shown)

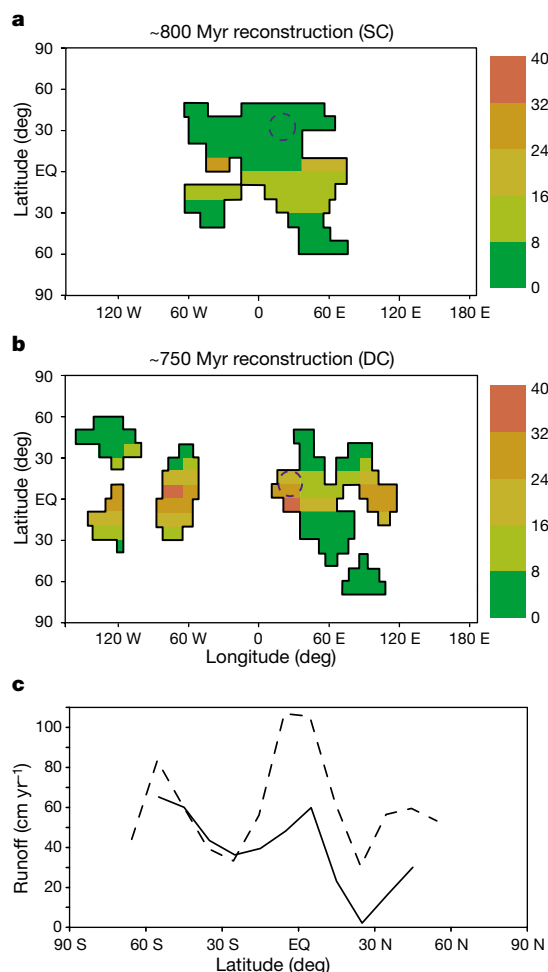


Figure 1 Palaeogeographies and hydrologic cycle. Weathering rates (10⁴ mol km⁻² yr⁻¹) simulated by the GEOCLIM model with a p_{CO_2} of 1,830 p.p.m.: **a**, for the continental reconstruction at ~800 Myr (denoted SC, supercontinent configuration), that is, before the Rodinia break-up, and **b**, for the continental reconstruction at ~750 Myr ago (denoted DC, dispersed configuration). Dashed circles, location of the basaltic provinces used for the runs including them. Accounting for the age error limits on these reconstructions, drift rates are between 13 and 8 cm yr⁻¹. Given that the break-up has been attributed to extensive volcanism and perhaps superplumes, these rates are quite consistent with the thermal regime proposed for the upper mantle. **c**, Annual mean zonal runoff (cm yr⁻¹) as simulated by the CLIMBER-2 model at a p_{CO_2} of 1,830 p.p.m. for the SC experiment (solid line) and for the DC experiment (dashed line).

and the global mean temperature is reduced to only 2 °C. Figure 1a, b shows the spatially-resolved weathering rates at 1,830 p.p.m. for the SC and DC experiments.

The vast size of the supercontinent restricts the delivery of precipitation from oceanic moisture sources (Fig. 1c). Thus, the SC experiment yields the driest continental climates and the lowest weathering rates. In contrast, the DC experiment is characterized by high runoff and large weathering rates in the equatorial regions, because the continents are smaller and widely dispersed with numerous oceans and seaways available as moisture sources. Thus, the larger runoff of the DC simulation results in higher consumption of CO₂ through weathering, and forces the climate to cool until the silicate weathering rate reaches a value equal to the SC simulation (itself equal to the degassing rate). Thus, using a complex model and quantifying the effect of the break-up, we predict a marked reduction in the concentration of CO₂ in the atmosphere to a persistently low value in the range 400–630 p.p.m. on timescales of >10 Myr (Fig. 2). We also find that model-predicted DC CO₂ values are in the range of radiative forcings, resulting in the build-up of ice sheets at latitudes greater than 30° (refs 5, 6). These CO₂ concentrations are just above the threshold value required to trigger a snowball Earth with the GEOCLIM model, that is, 250 p.p.m. (Fig. 2). Although a dispersed continental configuration is a common feature for reconstructions at 750 Myr ago, the true palaeogeography evolves during the snowball intervals via normal continental motion. Additional experiments (not shown) were performed to determine the sensitivity of our results to alternative dispersed configuration (for instance, the position of Siberia in Fig. 10b of ref. 20). Those runs demonstrate that the break-up effect may be larger (the atmospheric CO₂ level could be as low as 420 p.p.m. instead of 510 p.p.m. for the standard runs).

Although these experiments show that continental break-up could prepare the Earth for full glaciation, all the events that occurred before and during the dislocation have not been accounted for. Indeed, the break-up of Rodinia was heralded by intense volcanism, including the emplacement of large basaltic provinces between 825 and 750 Myr ago¹¹. Figure 3 illustrates the combined

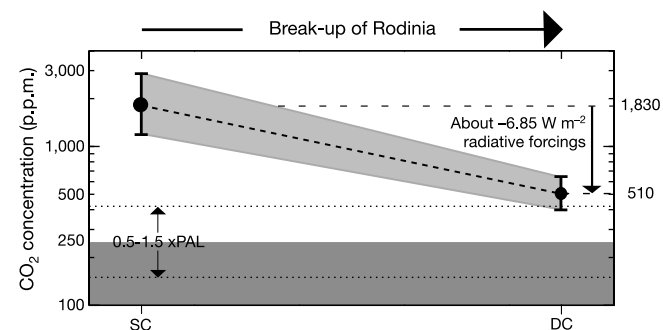


Figure 2 Effect of the Rodinia break-up on the greenhouse effect. Steady-state atmospheric CO₂ level achieved by the GEOCLIM model with the SC and DC reconstructions (Fig. 1). Vertical bars on the two data points denote the upper and the lower range of atmospheric CO₂ levels calculated using a 20% increase and a 20% decrease in degassing flux, respectively (see Methods). The horizontal dashed lines denote the exact CO₂ values reached by the model for the standard runs. The vertical arrow displays the change of the radiative forcing from direct CO₂ effects alone. The dark grey area shows the CO₂ levels required for a globally glaciated state in the 750-Myr-old climate simulations we used here. The dotted lines denote the range of atmospheric CO₂ levels required for low-latitude glaciations as predicted by a diverse array of climate models^{5,6} and are here for comparison with the threshold obtained by the GEOCLIM model. This range of CO₂ values is the consequence of several mechanisms that cannot be accounted for in a unique climate model. Among those mechanisms, ocean dynamics and sea-ice meltwater effects^{18,30} (but also the inclusion of an ice-sheet model⁵) can have a profound effect on the position of the critical collapse point. PAL, present atmospheric level.

effect of the weathering of freshly erupted basaltic provinces and continental separation. This increase in continental weatherability does not initially result in a significant change in the CO_2 levels as the basaltic surfaces are located in the sub-tropics (see the dashed circle, Fig. 1a), the driest regions in our simulations (Fig. 1c). Subsequently, the southward drift of the Laurentia plate (and also of the basaltic surfaces, Fig. 1b) and the dispersion of the continents increase the weathering rates and lower the CO_2 levels past the threshold required to trigger a snowball Earth (that is, 250 p.p.m., Fig. 3). Hence, our model results provide an explanation for the severe cooling observed during the Sturtian interval, and highlight the complex interplay between atmospheric CO_2 content, the dispersal of Rodinia, and the silicate/basaltic weathering feedback as the primary causes of the 'greenhouse'–'icehouse' climate transition.

Although we are now able to explain the onset of a snowball glaciation during the Sturtian glacial period, it is worth seeing if this model can be applied to the younger Neoproterozoic glacial episodes. Of these, the Marinoan glaciation (~650–600 Myr ago) is thought to be global in extent² whereas the global nature of the younger glacial pulse (Varangian, ~580–565 Myr ago) is questioned^{21–23}. Our present study indicates that dispersed continents positioned at low latitudes trigger reduction of global temperatures. Whether a continental configuration similar to the Sturtian one still applies to the younger glaciations is a matter of debate, primarily because of the lack of well-constrained palaeomagnetic data for the interval 700–580 Myr ago^{19,20}. At present, the palaeogeographic models can be divided into two categories. The first is a super-continent assembly, stretching from high southerly to low latitudes²⁰. The second posits a situation very similar to the Sturtian epoch, with most continents in mid to low latitudes²⁴. Although recent palaeomagnetic data and geochronologic data make a stronger case for the mid-to-low latitude configuration^{24,25}, these studies did not provide any global reconstructions that we could use in this investigation. However, we do note that if the low-to-mid latitude configuration is validated by further studies, then the similarities to the Sturtian situation suggest that our model may be applicable to the Marinoan event. We also note that the final trigger for this event was probably not basaltic surface weathering, but rather a transient methane-induced greenhouse, resulting in enhanced atmospheric CO_2 consumption⁷. This may explain the differences in the $\delta^{13}\text{C}$ excursions (both magnitude and timing) observed before both glaciations. Indeed, the available data suggest that the Sturtian glaciation was preceded by positive carbon isotopic values whereas

the Marinoan glaciation was preceded by an impressive but gradual decline in $\delta^{13}\text{C} > 10\text{‰}$ (ref. 23).

The latest Varangian event (the Gaskiers glaciation) was attributed to a regionally extensive glaciation rather than to a snowball-like glaciation^{1,21,23}, occurring during the pan-African orogen. We recently demonstrated that the CO_2 threshold required to initiate a snowball Earth when using the low-to-high latitude continental configuration proposed in ref. 20 is extremely low, around 80 p.p.m. (ref. 26). Furthermore, using the GEOCLIM model, we showed that this palaeogeography leads to restricted CO_2 consumption by silicate weathering (owing to the large percentage of continents located at high latitude), resulting in high steady-state CO_2 (around 1,500 p.p.m.)²⁶. Under this atmospheric CO_2 level, GEOCLIM simulates a snow accumulation pattern compatible with a regional scale glaciation occurring over the West Gondwana and the Laurentia cratons, and appears consistent with interpretations made from Varangian data^{1,21,23}.

In summary, this study provides a well-quantified mechanism to explain the Sturtian glaciation, which was perhaps the most extensive glaciation in the past billion years. We also note that palaeo-reconstructions for both younger glacial epochs are highly contentious, and a rigorous application of our model must await further palaeomagnetic data for these events. Furthermore, our study should also be considered as a basis for re-investigating Phanerozoic p_{CO_2} evolution by explicitly taking into account the tectonic forcing factor. More explicit modelling will help to reveal the importance of the tectonic forcing over all the Earth's history. □

Methods

General design of the GEOCLIM model

The COMBINE model is an atmosphere (one reservoir)–ocean (five reservoirs) geochemical model originally coupled to an 1D EBM¹⁰. Hence, this model incorporates latitudinal variation in temperature, runoff and exposed land area. The improvement of the COMBINE model that we perform here mainly consists in adding the longitudinal dimension to calculate the chemical weathering rates as the climatic variables (temperature and runoff) will now be provided by the CLIMBER-2 model on a 7×18 grid instead of the 1×18 grid of the EBM. The EBM is thus completely removed from the COMBINE code, and is replaced by CLIMBER-2 output. Moreover, the calculated runoff is no longer the result of a parametric law involving the zonally averaged continental surface and air temperature, but is now calculated by the CLIMBER-2 model, which explicitly simulates the hydrologic cycle.

Indirect coupling

A full coupling between COMBINE and CLIMBER-2 cannot be achieved owing to excessive computation times. For both continental configurations, shared boundary conditions include a solar luminosity 6% below the present-day value. In addition, we assume that the Earth's orbit around the Sun was circular (eccentricity = 0) and that the Earth's obliquity was 23.5° . This setting leads to an equal annual insolation for both hemispheres. The land surface in the climate model has the radiative characteristics of a desert and has a uniform elevation of 100 m for both palaeogeographies. Otherwise, the climate model was run until equilibrium (5,000 yr) under an atmospheric CO_2 concentration of 10,000 p.p.m. as an initial condition. Then starting from the previous equilibrium state each time, we conducted a set of simulations (5,000 yr for each to reach equilibrium) in which CO_2 levels were decreased step by step until we reach the CO_2 level for which a global glaciation occurred. More than 40 simulations have been performed in this way in order to have, at least, a climate simulation for each 1°C of global cooling.

We assume a linear behaviour in between each climatic simulation and thus, by doing a linear interpolation, we obtain the climatic variables of interest (temperature and runoff) for any p_{CO_2} on a 7×18 grid. These climatic parameters allow the calculation of the weathering rates within the 126 grid elements using the weathering law updated in the ref. 13. Fixing the CO_2 degassing to a given constant value, COMBINE is run until a steady-state p_{CO_2} is reached. At each time step of the calculation, temperature and runoff for each grid element are calculated through linear interpolation of the CLIMBER-2 results, for the coeval p_{CO_2} calculated by COMBINE. Because all results are displayed once carbon cycle steady-state is reached (after 4 to 6 simulated million years), the exact kinetics used to describe the various deposition processes as well as the weathering of continental oxidized and reduced sedimentary carbon are not of primary importance for this study, influencing only the response time and behaviour of the model during the steady-state approach. The only two parameters totally defining the steady-state p_{CO_2} are the volcanic CO_2 degassing, and the global continental silicate weathering summed up on each grid cell². An exhaustive description of the COMBINE model is thus not of primary importance for the present study, but can be found in ref. 10.

Regarding the volcanic CO_2 degassing, it might be argued that the break-up of a supercontinent may be characterized by increasing CO_2 degassing flux (via rift-related volcanism), counteracting the global cooling initiated by enhanced weathering. However,

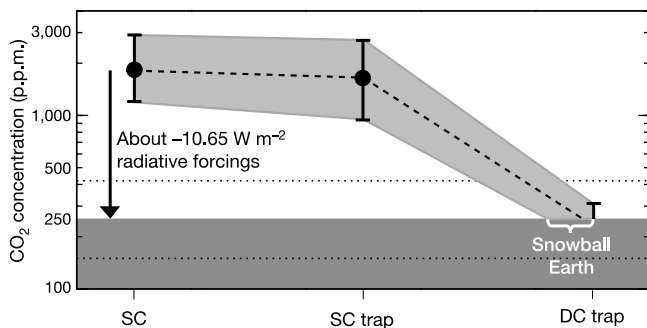


Figure 3 Atmospheric CO_2 history during the time period preceding the Sturtian snowball event. Steady-state atmospheric CO_2 level achieved by the GEOCLIM model for the SC runs, for the SC runs with the inclusion of basaltic provinces (denoted SC trap), and for the DC runs also with basaltic provinces (denoted DC trap). Vertical bars denote the upper and the lower range of atmospheric CO_2 levels calculated using (for the SC runs) a 20% increase and a 20% decrease in degassing flux, and (for the SC trap and DC trap runs) a 20% increase and a basaltic surface of $4 \times 10^6 \text{ km}^2$ and a 20% decrease and a basaltic surface of $8 \times 10^6 \text{ km}^2$ (see Methods). Dark grey area as Fig. 2. The vertical arrow displays the change of the radiative forcing from direct CO_2 effects alone.

despite the fact that degassing probably increased during the early phase of continental break-up, there is no clear evidence for long-term (10^7 -yr timescale) sustained degassing increase. Indeed, no clear consensus exists between the various reconstructions of the degassing flux through the geological past. For instance, recent re-evaluation of ridge production since 180 Myr ago suggests a roughly constant degassing rate since the middle Jurassic²⁷, in complete disagreement with previous reconstruction²⁸. To account for the hypothetical changes in the background degassing flux, we run the model for each continental configuration with values of the degassing flux ranging from a 20% decrease to a 20% increase relative to the present-day value (6.8×10^{12} mol of carbon per year; this is the value required to balance the global consumption through the weathering of silicate lithologies²⁹). These choices are somewhat arbitrary, but serve as a reasonable basis for probing the sensitivity of the system (see Figs 2 and 3).

Regarding the global weathering of continental surfaces, we use the most recent field weathering law for granitic lithology¹³ to calculate this flux as a function of the air temperature and continental runoff for each grid element. The relative proportion of silicate and carbonate outcrops is assumed to be the same in each grid cell because of the lack of precise lithological control. For the runs accounting for the basaltic provinces, the exact locations of the basaltic surface and the supports for such emplacement are fully described in ref. 8 and are broadly shown in Fig. 1. The total basaltic surface is 6×10^6 km² for the standard runs.

Received 7 November 2003; accepted 6 February 2004; doi:10.1038/nature02408.

- Evans, D. Stratigraphic, geochronological, and paleomagnetic constraints upon the Neoproterozoic climatic paradox. *Am. J. Sci.* **300**, 347–433 (2000).
- Sohl, L. E., Christie-Blick, N. & Kent, D. V. Paleomagnetic polarity reversals in Marinoan (ca 600 Ma) glacial deposits of Australia: Implications for the duration of low-latitude glaciation in Neoproterozoic time. *Geol. Soc. Am. Bull.* **111**, 1120–1139 (1999).
- Kirschvink, J. L. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C. C.) 51–52 (Cambridge Univ. Press, Cambridge, 1992).
- Hoffman, P. F. & Schrag, D. P. The snowball Earth hypothesis: testing the limits of global change. *Terra Nova* **14**, 129–155 (2002).
- Hyde, W. T., Crowley, T. J., Baum, S. K. & Peltier, R. W. Neoproterozoic 'snowball Earth' simulations with a coupled climate/ice-sheet model. *Nature* **405**, 425–429 (2000).
- Donnadieu, Y., Fluteau, F., Ramstein, G., Ritz, C. & Besse, J. Is there a conflict between the Neoproterozoic glacial deposits and the snowball Earth model: an improved understanding with numerical modelings. *Earth Planet. Sci. Lett.* **208**, 101–112 (2003).
- Schrag, D. P., Berner, R. A., Hoffman, P. F. & Halverson, G. P. On the initiation of a snowball Earth. *Geochim. Geophys. Res.* **3**, 101029/2001GC000219 (2002).
- Goddéris, Y. *et al.* The Sturtian "snowball" glaciation: fire and ice. *Earth Planet. Sci. Lett.* **211**, 1–12 (2003).
- Petoukhov, V. *et al.* CLIMBER-2: a climate system model of intermediate complexity. Part I: model description and performance for present climate. *Clim. Dyn.* **16**, 1–17 (2000).
- Goddéris, Y. & Joachimski, M. M. Global change in the late Devonian: modelling the Frasnian–Famennian short-term carbon isotope excursions. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **202**, 309–329 (2004).
- Li, Z. X. *et al.* Geochronology of Neoproterozoic syn-rift magmatism in the Yangtze Craton, South China and correlations with other continents: evidence for a mantle superplume that broke up Rodinia. *Precamb. Res.* **22**, 85–109 (2003).
- Walker, J. C. G., Hays, P. B. & Kasting, J. F. A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. *J. Geophys. Res.* **86**, 9776–9782 (1981).
- Oliva, P., Viers, J. & Dupré, B. Chemical weathering in granitic crystalline environments. *Chem. Geol.* **202**, 225–256 (2003).
- Berner, R. A. The rise of plants and their effect on weathering and atmospheric CO₂. *Science* **276**, 544–546 (1997).
- Millot, R., Gaillardet, J., Dupré, B. & Allègre, C. J. The global control of silicate weathering rates and the coupling with physical erosion: new insights from rivers of the Canadian Shield. *Earth Planet. Sci. Lett.* **196**, 83–98 (2002).
- Dessert, C. *et al.* Erosion of Deccan Traps determined by river geochemistry: impact on the global climate and the ⁸⁷Sr/⁸⁶Sr ratio of seawater. *Earth Planet. Sci. Lett.* **188**, 459–474 (2001).
- Ganopolski, A., Rahmstorf, S., Petoukhov, V. & Claussen, M. Simulation of modern and glacial climates with a coupled model of intermediate complexity. *Nature* **391**, 351–356 (1998).
- Donnadieu, Y., Ramstein, G., Fluteau, F., Roche, D. & Ganopolski, A. The impact of atmospheric and oceanic heat transports on the sea ice-albedo instability during the Neoproterozoic. *Clim. Dyn.* (in the press).
- Meert, J. G. A synopsis of events related to the assembly of eastern Gondwana. *Tectonophysics* **362**, 1–40 (2003).
- Meert, J. G. & Torsvik, T. H. The making and unmaking of a supercontinent: Rodinia revisited. *Tectonophysics* **375**, 261–288 (2003).
- Barfod, G. H. *et al.* New Lu–Hf and Pb–Pb constraints on the earliest animal fossils. *Earth Planet. Sci. Lett.* **201**, 203–212 (2002).
- Knoll, A. H. Learning to tell Neoproterozoic time. *Precamb. Res.* **100**, 3–20 (2000).
- Rice, A. H. N., Halverson, G. P. & Hoffman, P. F. Three for the Neoproterozoic: Sturtian, Marinoan and Varangerian glaciations. *Geophys. Res. Abstr.* **5**, 11425 (2003).
- Pisarevsky, S. A., Komissarova, R. A. & Khranov, A. N. New palaeomagnetic result from Vendian red sediments in Cisbaikalia and the problem of the relationship of Siberia and Laurentia in the Vendian. *Geophys. J. Int.* **140**, 598–610 (2000).
- Trindade, R. I. F., Font, E., D'Agrella-Filho, M. S., Nogueira, A. C. R. & Riccomini, C. Low-latitude and multiple geomagnetic reversals in the Neoproterozoic cap carbonate, Amazon craton. *Terra Nova* **15**, 441–446 (2003).
- Donnadieu, Y., Goddéris, Y., Ramstein, G. & Fluteau, F. in *Multidisciplinary Studies Exploring Extreme Proterozoic Environment Conditions* (eds Jenkins, G. S., McMenamin, M., McKay, C. P. & Sohl, L. E.) (American Geophysical Union, Washington DC, in the press).
- Rowley, D. B. Rate of plate creation and destruction: 180 Ma to present. *Geol. Soc. Am. Bull.* **114**, 927–933 (2002).

- Gaffin, S. Ridge volume dependence on seafloor generation rate and inversion using long term sea level change. *Am. J. Sci.* **287**, 596–611 (1987).
- Gaillardet, J., Dupré, B., Louvat, P. & Allègre, C. J. Global silicate weathering and CO₂ consumption rates deduced from the chemistry of the large rivers. *Chem. Geol.* **159**, 3–30 (1999).
- Poulsen, C. J., Pierrehumbert, R. T. & Jacob, R. L. Impact of ocean dynamics on the simulation of the Neoproterozoic "snowball Earth". *Geophys. Res. Lett.* **28**, 1575–1578 (2001).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.D. (tiphe@lsce.saclay.cea.fr).

Millennial and orbital variations of El Niño/Southern Oscillation and high-latitude climate in the last glacial period

Chris S. M. Turney^{1*}, A. Peter Kershaw², Steven C. Clemens³, Nick Branch⁴, Patrick T. Moss⁵ & L. Keith Fifield⁶

¹School of Archaeology and Palaeoecology, Queen's University, Belfast BT7 1NN, UK

²School of Geography and Environmental Science, Monash University, Victoria 3800, Australia

³Geological Sciences, Brown University, Providence, Rhode Island 02912-1846, USA

⁴Department of Geography, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK

⁵Department of Geography, University of Wisconsin-Madison, 550 N Park Street, Madison, Wisconsin 53706, USA

⁶Department of Nuclear Physics, Research School of Physical Sciences and Engineering, Australian National University, Canberra, Australian Capital Territory 0200, Australia

* Present address: School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia

The El Niño/Southern Oscillation (ENSO) phenomenon is believed to have operated continuously over the last glacial–interglacial cycle¹. ENSO variability has been suggested to be linked to millennial-scale oscillations in North Atlantic climate during that time^{2,3}, but the proposals disagree on whether increased frequency of El Niño events, the warm phase of ENSO, was linked to North Atlantic warm or cold periods. Here we present a high-resolution record of surface moisture, based on the degree of peat humification and the ratio of sedges to grass, from northern Queensland, Australia, covering the past 45,000 yr. We observe millennial-scale dry periods, indicating periods of frequent El Niño events (summer precipitation declines in El Niño years in northeastern Australia). We find that these dry periods are correlated to the Dansgaard–Oeschger events—millennial-scale warm events in the North Atlantic climate record—although no direct atmospheric connection from the North Atlantic to our site can be invoked. Additionally, we find climatic cycles at a semiprecessional timescale (~11,900 yr). We suggest that climate variations in the tropical Pacific Ocean on millennial as well as orbital timescales, which determined precipitation in northeastern Australia, also exerted an influence on North Atlantic climate through atmospheric and oceanic teleconnections.

Within the North Atlantic region, the last glacial period (60–10 kyr ago) was characterized by a series of rapid millennial-scale climatic oscillations referred to as Dansgaard–Oeschger (D–O) events that have been identified throughout the Northern Hemi-

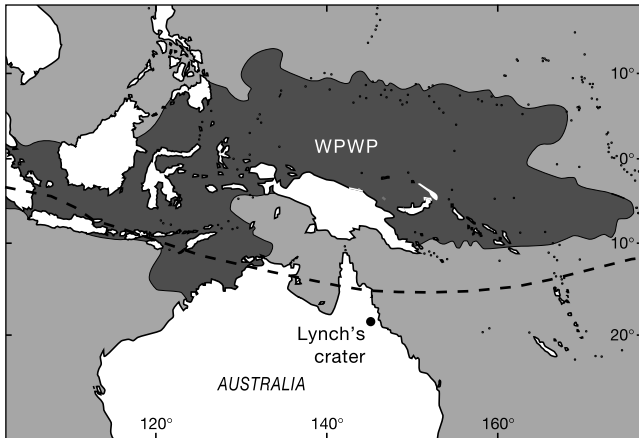


Figure 1 Location of Lynch's crater relative to the West Pacific Warm Pool (WPWP). The mean position of the intertropical convergence zone (ITCZ) during the austral summer (January) is shown as a dashed line. Modified from ref. 6.

sphere⁴. These D–O events can be bundled into regular, decreasing amplitude, cooling cycles ('Bond' cycles⁴) that culminate in massive discharges of ice into the North Atlantic (Heinrich or H-events) originating from either the Laurentide (H-1, -2, -4 and -5)⁴ or Fennoscandian (H-3 and -6) ice sheets⁵. Two models have been proposed to account for these extreme shifts in climate, namely a reorganization of the ocean's thermohaline circulation or changes in the dynamics of the tropical atmosphere–ocean system (for example, ENSO)².

Research has been directed towards determining whether ENSO was active before the Holocene epoch, and if so, determining the

amplitude and frequency of events^{1,3,6}. Multidecadal sequences from marine corals through the past 130,000 yr indicate that ENSO was in operation throughout this time¹. It has been suggested that the tropical ocean may lead high-latitude warming during the last glacial period⁷, though this relationship is contradicted by evidence for elevated West Pacific ocean surface salinity during stadials³. Concerns regarding the choice of foraminiferal species⁸, and the potential teleconnection to the North Atlantic through the East Asian monsoon system⁹, make the latter interpretation uncertain, however. Moreover, contemporary observations¹⁰ and modelling experiments¹¹ suggest that ENSO events increase temperatures/precipitation across the North Atlantic on interannual and millennial timescales.

The long, continuous terrestrial record from Lynch's crater (17° 37' S, 145° 70' E) provides the main reference for late Quaternary environmental change in northeastern Australia (Fig. 1)^{12,13}. The site is extremely sensitive to changes in the strength of ENSO activity, is located along steep precipitation and vegetation gradients, and lies to the south of the West Pacific Warm Pool (WPWP). Under contemporary conditions, approximately 80% of the annual rainfall (2,500 mm) falls during the austral summer months (November to March), mostly related to delivery by the southeast trade winds when the intertropical convergence zone (ITCZ) is at its most southerly extent¹⁴. During El Niño episodes, a northward movement of the ITCZ and a northeastward migration of the South Pacific convergence zone results in a significant decrease in summer precipitation (typically 150–300 mm below seasonal average) over the region^{14,15}. Although of importance today over more northerly latitudes in Australia¹⁴, the summer monsoon delivers little moisture to the Lynch's crater region and did not operate effectively during the last glacial period owing to the relatively northerly displaced ITCZ and the greatly increased Sunda and Sahul continental shelves^{16–18}, reducing cross-equatorial flow and effectively

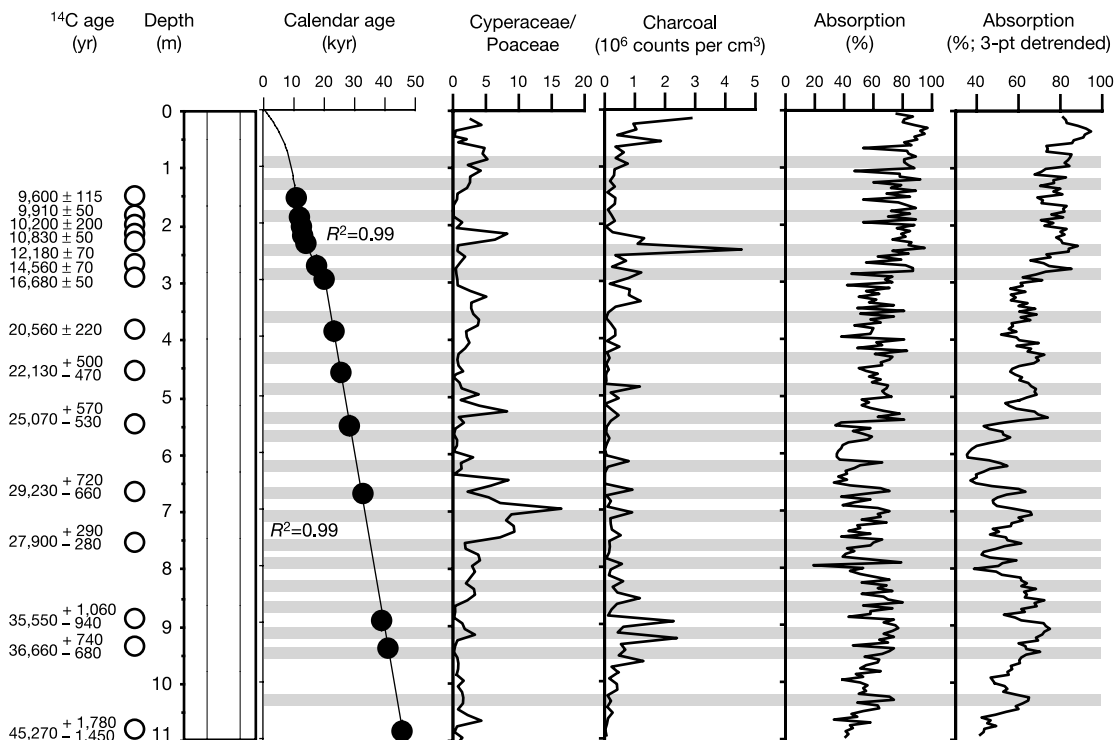


Figure 2 Summary of palaeoenvironmental data from Lynch's crater, northern Queensland, Australia. Radiocarbon ages <20,000 yr have been calibrated using the INTCAL98 interactive program (<http://depts.washington.edu/qil/calib/calib.html>), while greater radiocarbon ages were calibrated against the Nordic Sea record^{21,23}. (Radiocarbon

ages are given $\pm 1\sigma$.) The calibrated ages indicate two different phases of sedimentation, best represented by a third-order polynomial from the surface to 2.9 m, and a linear sedimentation rate from 2.9 to 11 m. Peat sediments make up the uppermost 11 mm of the sequence. Grey zonation denotes inferred dry periods.

decoupling northern Australia from the East Asian monsoon regime (and any associated North Atlantic atmospheric teleconnection). In contrast, the trade winds would have remained significant owing to the relatively narrow continental shelf off northeastern Australia. Consequently, Lynch's crater is ideally located to explore changes in past ENSO, independent of the North Atlantic region.

Peat sediments, which make up the upper 11 m of the record, were subjected to multi-proxy analysis (Fig. 2). Here we use the ratio of Cyperaceae (sedge) to Poaceae (grass) as a vegetational response to wetting (high ratio) and drying (low ratio) in the terrestrial sclerophyll and swamp surfaces on and around Lynch's crater. Charcoal counts were made alongside pollen analyses with high values considered to represent periods of increased burning¹⁹. As a proxy of surface wetness at the time of peat deposition, we used the degree of peat decomposition (humification) measured by optical absorption²⁰. Using a sampling resolution twice that of the other proxies, high values of absorption are interpreted as reflecting dry surface conditions, because increased microbial activity under aerobic conditions increases the degree of peat humification²⁰. A chronology for the sequence is provided by 15 radiocarbon ages¹⁹ (Fig. 2). Calibration was undertaken using the Nordic Sea radiocarbon calibration curve²¹ that utilizes the GISP2 isotope chronology²² and has been recognized as the most reliable of the published data sets²³.

Throughout the past 45,000 yr there are clear oscillations in the proxies (Fig. 2). The ratio of Cyperaceae to Poaceae fluctuates significantly, indicating alternation between wetter and drier conditions. Troughs in the ratio appear to coincide with relatively high levels of charcoal concentration, indicating associated burning at these times. Humification analyses also clearly record a significant range in absorption values, with high-amplitude and frequent changes in absorption that appear to parallel the above trends. High values are indicative of greater humification, generally corresponding with high burning levels and a decreased Cyperaceae/Poaceae ratio, and the reverse with low absorption values.

Spectral analysis²⁴ was undertaken on the high-resolution peat humification data. The record was detrended, weighted, smoothed and interpolated to a constant 200-yr interval. Two significant spectral frequencies are readily discernible over the past 45,000 yr: a semiprecessional cycle centred on 11.9 kyr, and a millennial-scale cycle at 1,490 yr (Fig. 3a). The 1,490-yr period is interpreted as changes in precipitation associated with long-term changes in ENSO in the tropical Pacific Ocean region, through a disruption of the trade winds.

Within the error of the analysis, the above period matches the 1,470-yr cyclicity of ENSO reported for the Holocene in southern Ecuador²⁵ and similar periods have been suggested for the last glacial period in precipitation and vegetation across North America²⁶ as a result of palaeo-ENSO cycles. Furthermore, the 1,490-yr period matches that of the D-O events in the North Atlantic^{4,27}. Thus, both millennial-scale ENSO and high-latitude climate change appear to have operated on millennial-scale periodicities during at least the last glacial period.

To evaluate potential linkage between the 1,490-yr humification cycle and the 1,470-yr cycle from the North Atlantic region, we undertook a cross-spectrum between the Lynch's crater humification and the GISP2 $\delta^{18}\text{O}$ record²² (Fig. 3a). This analysis indicates that variance in the two records at the ~1,500-yr period is linearly correlated above the 80% confidence interval, and that, within the limitations of the two age models, the two records are in phase. A similar observation can be made of the coherence between the Cyperaceae/Poaceae ratio and the GISP2 $\delta^{18}\text{O}$ record, which is given in the Supplementary Information. Lynch's crater becomes dry at the same time as Greenland temperature increases. Schulz²⁷ found that the 1,470-yr cyclicity is tied directly to D-O interstadials 5, 6 and 7 in the Greenland $\delta^{18}\text{O}$ record. To assess this, we conducted wavelet analyses of both time series (Fig. 3b and c). Both records show

maximum amplitude of the 1,470-yr cycle over the same period between 35 and 25 kyr ago, further supporting a very close relationship between drying in northern Australia and warming in Greenland during interstadial events. In addition, both wavelet plots indicate the presence of a semiprecession cycle. The semiprecession cycle in the GISP2 record is comparatively weak but present none-

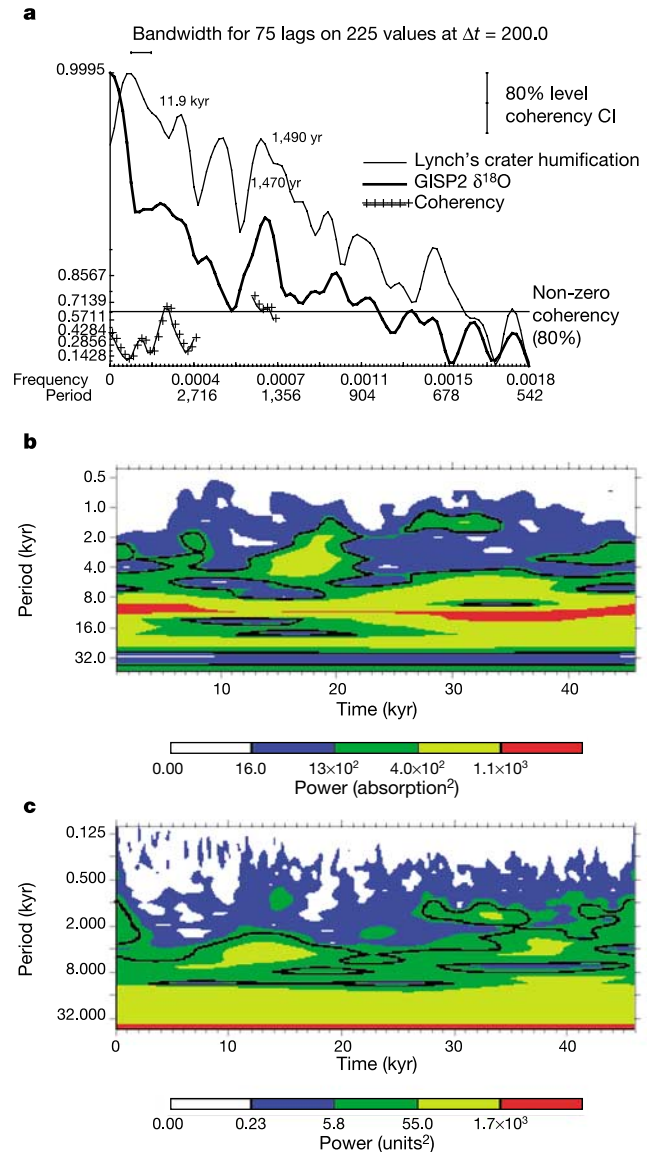


Figure 3 Cross-spectral and wavelet analysis of the Lynch's crater and GISP2 records for the past 50 kyr. **a**, Lynch's crater absorption and GISP2 spectrum using the Blackman-Tukey approach. The record was smoothed with a three-point gaussian filter (0.25, 0.5, 0.25), detrended using a weighted smooth, and interpolated to a constant 200-yr interval. Spectral and cross-spectral analysis of the proxy data sets was undertaken using the ARAND software package (<http://www.ngdc.noaa.gov/paleo/softlib/>)²⁴. Spectral density (s^2/f , where s^2 = variance and frequency f = 1/period) is normalized and plotted on a logarithmic scale. A higher-resolution (multi-taper; MT) spectral approach was also employed to provide another means of estimating the humification spectrum and judging the significance of spectral peaks. The periods (years) of spectral peaks (significant at the 95% confidence level) are labelled. The periods identified in the MT approach are consistent with the Blackman-Tukey results given the difference in resolution of the two methods. Also shown is the degree of coherence between the two data sets. **b**, **c**, Wavelet power spectrum documents the 1,490-yr and 11.9-kyr variability in ENSO for Lynch's crater (**b**) and GISP2 (**c**). The contour levels are selected so that 75%, 50%, 25% and 5% of the wavelet power is above each level, respectively. Analysis was undertaken using interactive software available at <http://paos.colorado.edu/research/wavelets/>.

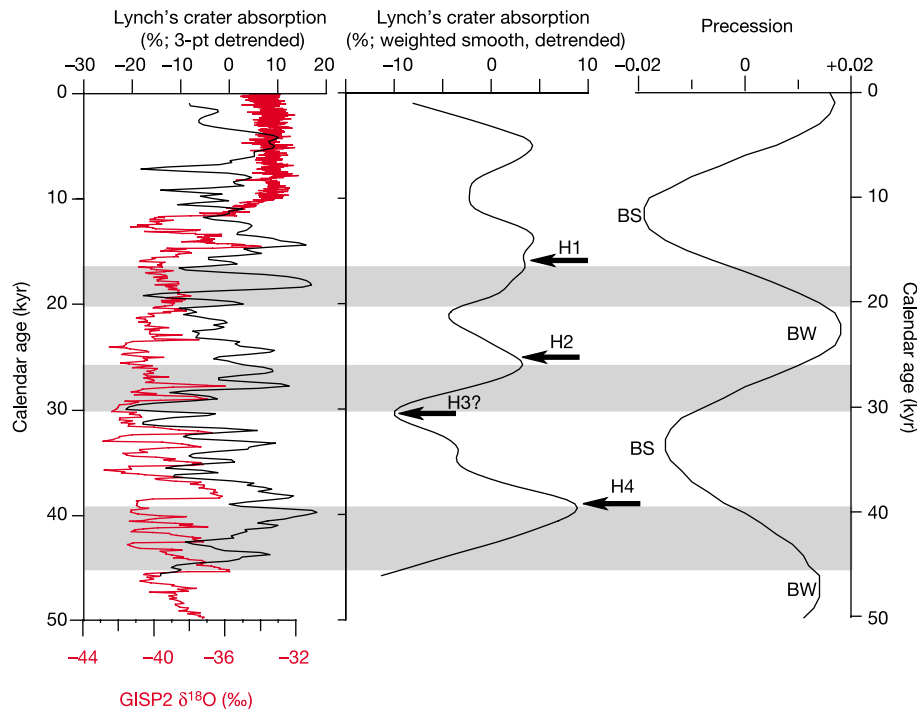


Figure 4 Comparison of Lynch's crater 1,490-yr and 11.9-kyr trends in absorption values, GISP2 $\delta^{18}\text{O}$ values²⁹ and the precessional component of orbital forcing¹. The 11.9-kyr curve was generated by applying a 15% weighted smooth to the detrended

absorption data. In the precessional curve, 'BS' and 'BW' denote the timing of the perihelion in the boreal summer and winter, respectively.

theless, whereas in Lynch's crater (Fig. 3b) the cycle is strong and present through the entire sequence (consistent with the dominance of the 11.9-kyr peak in Fig. 3a). Although we observe wetter conditions during stadials in the Lynch's crater data set (Fig. 4) (potentially consistent with increased cross-equatorial flow associated with a less intense summer East Asian monsoon^{3,9}), the semiprecessional cycle observed in the humification data set demonstrates longer-term changes in drying through the North Atlantic 'Bond' cycles. If the latter cooling trend were to increase cross-equatorial flow, we should also observe longer-term increases in wetness in the humid tropics. The opposite is recorded in Lynch's crater. We are therefore confident that, in contrast to earlier studies³, we are truly seeing cycles associated with long-term changes in ENSO that are independent of the monsoonal system (and by inference any link to the North Atlantic region). The data presented are consistent with semiprecessional changes in ENSO frequency in the tropical Pacific region with increasingly frequent 'warm' ENSO events (represented by drying in northern Australia), centred on 40, 25 and 15 kyr ago, while 'cold' events are centred on 30 and 21 kyr ago (Fig. 4).

Ocean-atmosphere modelling studies have proposed that the timing of the perihelion is a major control on long-term ENSO variability through asymmetric heating of the equatorial Pacific¹¹; periods of frequent El Niño events are expected during extreme strengths of the seasonal cycle on a semiprecessional cycle, that is, when the perihelion is in either the boreal spring or autumn. ENSO shutdown is predicted when perihelion is in either the boreal winter or summer. The Lynch's crater record lends support to this model, showing excellent agreement with the timing of more frequent 'warm' ENSO events on the semiprecessional timescale (Fig. 4). In addition, the three periods of increasing 'warm' ENSO activity identified here coincide with H-1, -2 and -4 in the GISP2 $\delta^{18}\text{O}$ data set. No such relationship is observed with H-3.

One explanation for the relationship between ENSO states, D-O and Heinrich events is that during an El Niño, warming on a northwest-southeast transect across North America takes place²⁶, causing potential melting of an ice front. An alternative explanation

for this relationship is increased Pacific sea surface temperatures associated with ENSO activity that would have led to a larger latitudinal temperature gradient, transporting more moisture to high latitudes^{28,29}, and ultimately creating a surge in ice³⁰. The absence of a long-term shift in ENSO activity on the semiprecessional timescale associated with H-3 indicates that a different mechanism must have played a role in the generation of high-latitude European ice discharge. In contrast, the changes induced by ENSO were sufficient to force changes in the mass balance of the Laurentide ice sheet. These changes led to a rapid melting or surging of the ice, and (through increased freshwater discharge) turned off North Atlantic deep water formation⁴, which led to hemispheric (and possibly global) climate change. We therefore consider the tropical Pacific Ocean to be a crucial component of the Earth-ocean-atmosphere system, operating on interannual, millennial and semiprecessional timescales. □

Received 29 July 2003; accepted 3 February 2004; doi:10.1038/nature02386.

1. Tudhope, A. W. *et al.* Variability in the El Niño-Southern Oscillation through a glacial-interglacial cycle. *Science* **291**, 1511–1517 (2001).
2. Broecker, W. S. Does the trigger for abrupt climate change reside in the ocean or in the atmosphere? *Science* **300**, 1519–1522 (2003).
3. Stott, L., Poulson, C., Lund, S. & Thunell, R. Super ENSO and global climate oscillations at millennial time scales. *Science* **297**, 222–226 (2002).
4. Bond, G. *et al.* Correlations between climate records from north Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
5. Grousset, F. E. *et al.* Patterns of ice-rafted detritus in the glacial North Atlantic (40–55°N). *Paleoceanography* **8**, 175–192 (1993).
6. Gagan, M. K. *et al.* New views of tropical paleoclimates from corals. *Quat. Sci. Rev.* **19**, 45–64 (2000).
7. Hendy, I. L. & Kennett, J. P. Tropical forcing of North Pacific intermediate water distribution during Late Quaternary rapid climate change? *Quat. Sci. Rev.* **22**, 673–689 (2003).
8. Spero, H. J., Mielke, K. M., Kalve, E. M., Lea, D. W. & Pak, D. K. Multispecies approach to reconstructing eastern equatorial Pacific thermocline hydrography during the past 360 kyr. *Paleoceanography* **18**, 1–16 (2003).
9. Wang, Y. J. *et al.* A high-resolution absolute-dated late Pleistocene monsoon record from Hulu Cave, China. *Science* **294**, 2345–2348 (2001).
10. Hoerling, M. & Kumar, A. The perfect ocean for drought. *Science* **299**, 691–694 (2003).
11. Clement, A. C., Cane, M. A. & Seager, R. An orbitally driven tropical source for abrupt climate change. *J. Clim.* **14**, 2369–2375 (2001).
12. Kershaw, A. P. Record of last interglacial-glacial cycle from northeastern Queensland. *Nature* **72**, 159–161 (1978).

13. McGlone, M. S., Kershaw, A. P. & Markgraf, V. in *El Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation* (eds Diaz, H. F. & Markgraf, V.) 435–462 (Cambridge Univ. Press, Cambridge, 1992).
14. Godfred-Spenning, C. R. & Reason, C. J. C. Interannual variability of lower-tropospheric moisture transport during the Australian Monsoon. *Int. J. Clim.* **22**, 509–532 (2002).
15. Dai, A. & Wigley, T. M. L. Global patterns of ENSO-induced precipitation. *Geophys. Res. Lett.* **27**, 1283–1286 (2000).
16. Roggon, P. & Williams, M. A. J. Late Quaternary climatic changes in Australia and North Africa: A preliminary interpretation. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **21**, 285–327 (1977).
17. Nanson, G. C., Price, D. M. & Short, S. A. Wetting and drying of Australia over the past 300 ka. *Geology* **20**, 791–794 (1992).
18. Johnson, B. J. *et al.* 65,000 years of vegetation change in central Australia and the Australian summer monsoon. *Science* **284**, 1150–1152 (1999).
19. Turney, C. S. M. *et al.* Redating the onset of burning at Lynch's crater (North Queensland): Implications for human settlement in Australia. *J. Quat. Sci.* **16**, 767–771 (2001).
20. Blackford, J. J. & Chambers, F. M. Determining the degree of peat decomposition for peat based palaeoclimatic studies. *Int. Peat J.* **5**, 7–24 (1993).
21. Voelker, A. H. L., Grootes, P. M., Naudeau, M.-J. & Sarinthein, M. Radiocarbon levels in the Iceland Sea from 25–53 kyr and their link to the Earth's magnetic field intensity. *Radiocarbon* **42**, 437–452 (2000).
22. Stuiver, M. & Grootes, P. M. GISP2 oxygen isotope ratios. *Quat. Res.* **53**, 277–284 (2000).
23. Laj, C. *et al.* Geomagnetic field intensity, North Atlantic Deep Water circulation and atmospheric $\Delta^{14}\text{C}$ during the last 50 kyr. *Earth Planet. Sci. Lett.* **200**, 177–190 (2002).
24. Jenkins, G. M. & Watts, D. G. *Spectral Analysis and its Applications* (Holden Day, Oakland, 1968).
25. Moy, C. M., Seltzer, G. O., Rodbell, D. T. & Anderson, D. M. Variability of El Niño/Southern Oscillation activity at millennial timescales during the Holocene epoch. *Nature* **420**, 162–165 (2002).
26. Wang, H., Follmer, L. R. & Liu, J. C. Isotope evidence of paleo-El Niño–Southern Oscillation cycles in loess–paleosol record in the central United States. *Geology* **28**, 771–774 (2000).
27. Schulz, M. The tempo of climate change during Dansgaard-Oeschger interstadials and its potential to affect the manifestation of the 1470-year climate cycle. *Geophys. Res. Lett.* **29**, doi:2001GL013277 (2002).
28. Kukla, G. J., Clement, A. C., Cane, M. A., Gavin, J. E. & Zebiak, S. E. Last interglacial and early glacial ENSO. *Quat. Res.* **58**, 27–31 (2002).
29. Rind, D. Latitudinal temperature gradients and climate change. *J. Geophys. Res.* **103**, 5943–5971 (1998).
30. MacAyeal, D. R. Binge/purge oscillations of the Laurentide Ice Sheet as a cause of the North Atlantic's Heinrich Events. *Eos* **74**, 359 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Gagan and B. Opdyke for suggestions about earlier drafts of this manuscript. C.S.M.T. acknowledges fellowships from the Australian Research Council and the Menzies Institute (King's College, London) held at the Australian National University. We thank S. Haberle and J. Luly for assistance in the field, G. Anderson and A. Crawford for permission to core Lynch's crater and E. Turton for laboratory preparations. This work was supported by The National Geographic, the ARC, Monash University, The Royal Society, ANSTO and NERC.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.S.M.T. (cturney@uow.edu.au or c.turney@qub.ac.uk).

Experimental evidence for apparent competition in a tropical forest food web

Rebecca J. Morris¹, Owen T. Lewis² & H. Charles J. Godfray¹

¹NERC Centre for Population Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK

²Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

The herbivorous insects of tropical forests constitute some of the most diverse communities of living organisms¹. For this reason it has been difficult to discover the degree to which these communities are structured, and by what processes. Interspecific competition for resources does occur, but its contemporary importance is limited because most pairs of potentially competing insects feed on different host plants². An alternative way in which species can interact is through shared natural enemies, a

process called apparent competition³. Despite extensive theoretical discussion there are few field demonstrations of apparent competition, and none in hyper-diverse tropical communities. Here, we experimentally removed two species of herbivore from a community of leaf-mining insects in a tropical forest. We predicted that other species that share natural enemies with the two removed species would experience lower parasitism and have higher population densities in treatment compared with control sites. In both cases (on removal of a dipteran and a coleopteran leaf-miner species) we found significantly lower parasitism, and in one case (removal of the dipteran) we found significantly higher abundance a year after the manipulation. Our results suggest that apparent competition may be important in structuring tropical insect communities.

Apparent competition is a type of indirect interaction^{4,5} that is defined as a negative effect of one species on the population growth rate or abundance of another species, mediated through the action of shared natural enemies. It can occur whether or not the two species compete directly for resources³. Because it can produce patterns in communities that resemble those caused by resource competition (hence its name), apparent competition has been invoked widely as a process that might structure assemblages of species in which ordinary competition is not pervasive⁶, for example for assemblages of herbivorous insects feeding on a diverse array of host plants⁷. The importance of apparent competition has been demonstrated experimentally in ecosystems such as the marine littoral^{8,9} and between vertebrates mediated by pathogens¹⁰, but the evidence from terrestrial invertebrate communities is less clear. Short-term apparent competition (responses within a generation) has been observed in temperate aphid communities^{11,12}, and long-term apparent competition (operating over more than one generation) has been demonstrated in laboratory systems¹³ and has been invoked to explain changes in the abundance of agricultural pests¹⁴. However, we lack evidence that apparent competition can lead to long-term changes in natural insect communities, and in particular whether it may be important in tropical rainforests, where parasitism and predation can be particularly intense¹⁵ and where levels of insect biodiversity are exceptionally high.

We investigated the role of apparent competition in structuring a community of leaf-mining insects in moist tropical forest in Belize, Central America. Leaf miners are a phylogenetically diverse group whose larvae feed inside the leaf lamina. They suffer severe mortality from a guild of parasitoid wasps that are restricted to leaf miners but which show differing degrees of host specificity¹⁶. Previously, we constructed a quantitative food web to describe the community and how it was interlinked¹⁷, and found that 93 species of leaf miner were attacked by 84 species of parasitoid wasp (Fig. 1). Quantitative food webs provide an indication of the likely occurrence of interactions involving apparent competition, but as food webs are static, an experimental approach is necessary to confirm its presence. We used the food web to identify leaf-miner species that had the potential to interact strongly with other species through apparent competition.

Of the plants that were host to leaf miners, most were attacked by a single species. *Lepidaploa tortuosa* (Asteraceae), however, was mined by two common species: a fly, *Calycomyza* sp. 8 (Diptera: Agromyzidae), and a beetle, *Pentispa fairmairei* (Coleoptera: Chrysomelidae). We tested a series of hypotheses drawn from apparent competition theory¹⁸: that removal of the fly would lead to lower parasitism and higher abundance of other dipteran leaf miners with which this species shares parasitoids on different host plants, and that removal of the beetle would have the same effect on another beetle, *Pachyschelus collaris* (Coleoptera: Buprestidae), whose only recorded parasitoid more commonly attacks *P. fairmairei* (Fig. 1). As described in the Methods, in December 2001 we manipulated the abundance of the two focal leaf-miner species by removing *L. tortuosa* in six replicate sites. Each treatment site was paired with a

Parasitoids
Scale: hosts $\times 4.53$

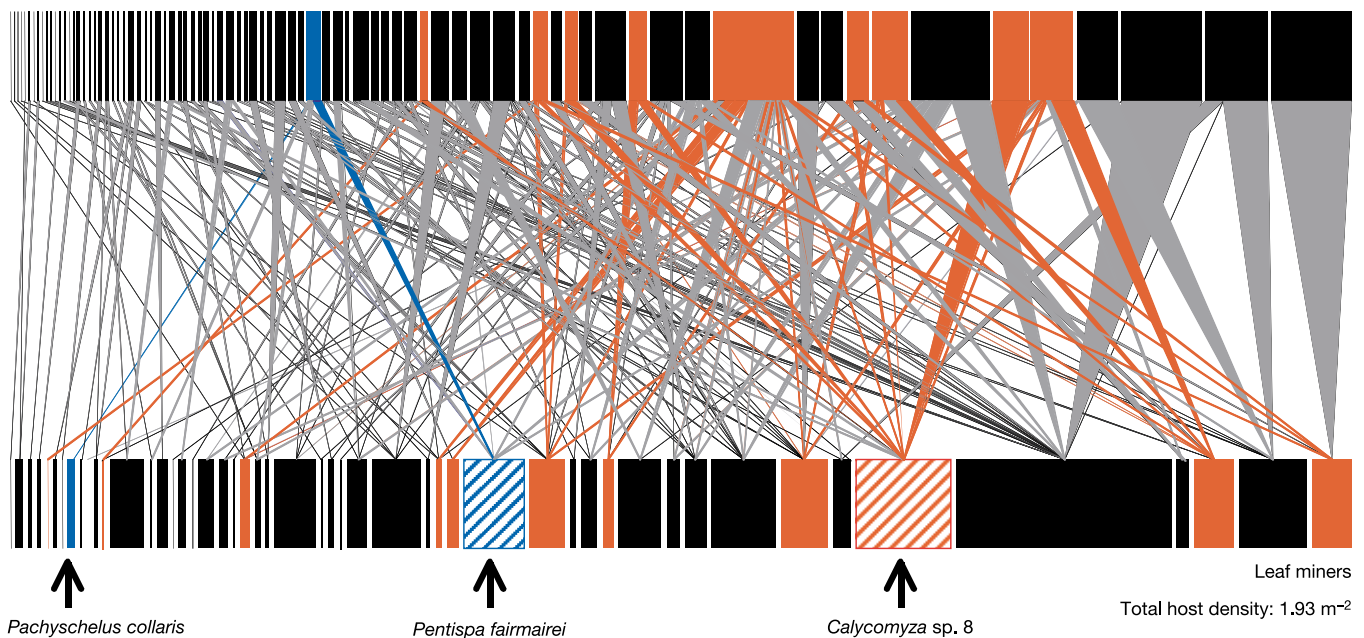


Figure 1 Quantitative food web¹⁷ showing leaf-miner species (bottom bars), parasitoid species (top bars), trophic links among them, and the species predicted to be affected by the manipulation. *Calycomyza* sp. 8 and *P. fairmairei* were directly affected by host plant removal. Dipteran leaf-miner species present during the sampling period and

predicted to be affected indirectly via parasitoids shared with *Calycomyza* sp. 8 are shown in red. The beetle *P. collaris* (blue) was also predicted to be affected indirectly by the manipulation through parasitoids that it shares with *P. fairmairei*. Only hosts from which parasitoids were reared are shown in the web.

control site in which a similar volume of foliage was removed, but from plants that were not attacked by leaf miners.

The abundances of the target species and the proportion parasitized were recorded approximately 5–6 leaf-miner generations after the manipulation, between October and December 2002. The dipteran leaf miners that shared parasitoids with *Calycomyza* sp. 8 together experienced 51% mortality due to parasitoid attack in control sites. As predicted, a significantly lower proportion was parasitized in treatment compared with control sites, with the manipulation reducing parasitism by a multiplicative factor of 0.6 (for details see Fig. 2a and Methods). The average density of the dipteran leaf miners was also higher in treatment compared with control sites (Fig. 2b). *Pachyschelus collaris* was at relatively low densities during the sampling period, and experienced 88% mortality due to parasitoid attack in control sites. As predicted, parasitism was lower in treatment compared with control sites, and although the estimated marginal multiplicative reduction was strong (0.1), this was associated with a large standard error (Fig. 2a). *Pachyschelus collaris* was not significantly more abundant in treatment sites, although there was a nonsignificant trend in this direction ($P = 0.14$; Fig. 2b). Thus the response of the dipteran leaf miners to the perturbation provides strong support for our predictions, whereas that of the coleopteran leaf miner is weaker but consistent.

We interpret our results as being due to indirect interactions between hosts, mediated by natural enemies. However, because the only way that we could practically manipulate leaf-miner densities was by removing the host plant, *L. tortuosa*, there is the possibility that indirect interactions between the plant and other herbivores might have influenced our results. Although we cannot formally exclude this (and such a result in itself would be interesting) we think it is unlikely. The plant constituted less than 0.01% of the biomass in the site and the manipulation caused minimal alteration to vegetation structure. We have no evidence that this plant provides important resources or is particularly attractive to parasitoids, and

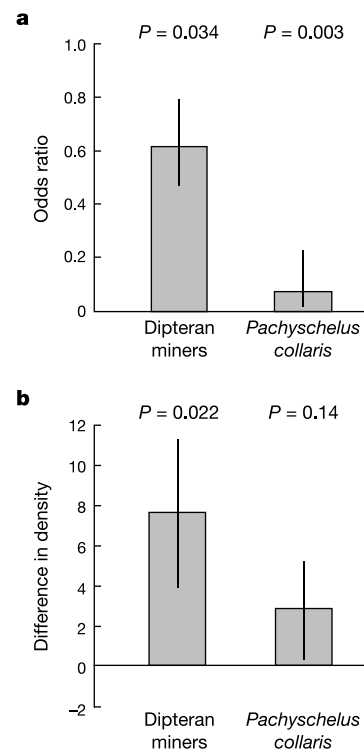


Figure 2 Effect of species removal on parasitism (a) and species abundances (b) 1 year after the manipulation. Differences in parasitism are expressed as odds ratios, the multiplicative effect of treatment, controlling for mean effects of plot and species. The null hypothesis is that the odds ratio equals one. Differences in abundance are expressed as the difference in density of leaf miners per plot caused by treatment, controlling for mean effects of plot and species. The null hypothesis is that there is no difference. Estimated effects with standard errors are shown for the two response groups. The probability associated with the test of the hypothesis that species removal will lead to lower parasitism and higher abundance is given above the histograms.

that this could explain why its absence reduces parasitism.

To our knowledge our results are the first demonstration of long-term apparent competition in natural communities of herbivorous insects, and they are also unusual in attempting to manipulate insect communities in the tropics. They support suggestions that interactions mediated by shared natural enemies may be a significant factor in structuring natural communities. Simple competition theory predicts that no two competitors can coexist on identical resources, and that competition can cause regularities along resource axes¹⁹. The equivalent apparent competition theory predicts that a shared natural enemy will drive all but one of its prey or hosts to extinction (leading to what has been called dynamic monophagy⁷), and that species can be distributed regularly over "natural-enemy space"²⁰. There is a large amount of literature investigating theoretically and experimentally the biological processes that permit competitors to coexist, but the equivalent body of work on the coexistence of apparent competitors is smaller and includes few experimental tests^{18,21–23}. Part of the reason for this is that, compared with competitive systems, the dynamic interactions between predators and prey, hosts and parasitoids, and so on, are much more frequently oscillatory and have a greater tendency to be unstable, making their analysis much harder²⁴. Field manipulations of host–parasitoid systems are also much more difficult than those of herbivore–plant systems. If our results prove to be typical of herbivore communities, then they suggest that the development of this theory, and its associated empirical tests, will be essential to understanding the diversity and structure of insect communities, especially in the species-rich tropics.

Our results also have implications for the study of biological perturbations. It is widely appreciated that introducing or removing a strong competitor or top predator can have major effects that cascade through a community²⁵. Much less attention has been paid to indirect effects mediated by natural enemies (but see refs 26, 27). For example, interventions such as selective logging may remove hosts and prey of natural enemies that also attack species feeding on other host plants. Similarly, introducing natural enemies, especially polyphagous species, may dynamically link hitherto weakly coupled parts of the community. This was highlighted recently by one study that constructed a quantitative parasitoid web based on Lepidoptera in an ostensibly pristine Hawaiian ecosystem and found the web to be dominated by polyphagous parasitoid species that had been introduced to control agricultural pests²⁸. Interventions such as selective logging and biological control are normally preferable to many of their environmentally more harmful alternatives, but a better understanding of apparent competition and related phenomena is required to ensure that their environmental impact is minimized. □

Methods

Study site

The field site was in moist tropical forest near the London Natural History Museum's Las Cuevas Research Station in the 170,000-ha Chiquibul Forest Reserve, Cayo District, southwest Belize. The forest is of relatively low stature because of periodic disruption by hurricanes. In previous work¹⁷ we characterized the quantitative structure of the community of leaf-mining insects, and their parasitoids, that attack the vegetation growing along the margins of cleared tracks at this site. Our study site was a 6-km stretch of track comprising a band of vegetation 3 m wide on each side of the 3-m wide track. It was divided into 12 plots, each 500 m in length, which were assigned alternately to treatment and control. At this site leaf miners occur at very low abundance in the forest canopy and interior understorey¹⁷ so the plots were likely to have experienced little immigration of leaf miners and parasitoids.

Manipulation

In treatment plots all individuals of the shrubby vine *L. tortuosa* were uprooted and removed in December 2001. Some re-growth occurred (we estimate about 5% of the original biomass) and the plant removal was repeated in May 2002. In control plots the same biomass of plant material was removed from plant species, chosen at random, that were not attacked by leaf miners. We chose this control so that the distribution of host plants with leaf miners was not affected. The manipulation might have had a transient direct effect on parasitoid abundance, but this is very unlikely to have influenced

parasitism rates by the time the data were collected (to minimize any effect we removed the plant at a time of year when its associated leaf miners were at low abundance). *Lepidaploa tortuosa* is host to the monophagous *Calycomyza* sp. 8, an agromyzid leaf miner that shares a group of specialist parasitoids with a series of other dipterans, chiefly in the genus *Calycomyza*, which we predicted would be influenced by its removal. *Lepidaploa tortuosa* is also host to two species of hispine chrysomelid, *P. fairmairei* and *Basiolus lineaticollis*, whose mines cannot be distinguished and are lumped together in the food web in Fig. 1. However, rearing the adults revealed that *B. lineaticollis* was uncommon, its mines constituting 10% of the total. *Pentispa fairmairei* was reared in small numbers from other Asteraceae species in the site, but *L. tortuosa* was its most important host. We predicted that perturbation of *P. fairmairei* would affect *P. collaris*, a buprestid that feeds on the herb *Desmodium canum* and shares all its recorded parasitoids with the former.

Sampling

The abundances and proportion parasitized of the target leaf-miner species were recorded between October and December 2002. To sample the site, each plot was divided into ten subplots, and then two subplots per plot (chosen randomly) were searched for leaf miners up to a height of 2 m, every 10–14 days until the whole site had been surveyed. Leaf miners were collected and reared in the laboratory to assess current population abundances and proportion parasitized^{17,29}. We checked that the manipulation had reduced the densities of *Calycomyza* sp. 8 and *P. fairmairei* by comparing the densities of all mines (both occupied and unoccupied) in control and treatment sites. For *Calycomyza* sp. 8 the effect of the manipulation was strongly significant ($F_{1,5} = 14.10$, $P = 0.0061$; log-linear generalized linear model adjusted for over-dispersion, one-tailed significance of treatment controlling for block) whereas for *P. fairmairei* the difference was more marginal ($F_{1,5} = 3.99$, $P = 0.05$; same statistical methods) because this species has alternative host plants.

Data analysis

Of the dipteran leaf-miner species that shared parasitoids with *Calycomyza* sp. 8, 12 were sufficiently common to contribute to a test of the abundance hypothesis; eight of these suffered parasitism in both treatment and control plots and thus could contribute to a test of the parasitism hypothesis. Ten of these dipteran species are highlighted in Fig. 1; a further two species were included in the analysis because additional data revealed that they also share parasitoids with *Calycomyza* sp. 8. Data were analysed using generalized linear modelling techniques implemented in the statistical package R (<http://www.r-project.org/>). For parasitism we used logistic regression, using the residual deviance to adjust standard errors if over-dispersion was present; the analysis of abundance assumed normally distributed errors. The effect of the treatment was assessed by adding it as a factor to a model already containing fixed plot effects and, in the case of the dipteran leaf miners, a fixed species factor. The test thus used a one-tailed F or χ^2 statistic depending on the type of regression and presence of over-dispersion. The effect of treatment on the proportion of leaf miners parasitized is illustrated as the estimated marginal multiplicative odds factor, and on abundance as the marginal increase in density. For the *Calycomyza* sp. 8 manipulation, the log odds effect of treatment on parasitism was -0.48 (standard error = 0.26, $F_{1,48} = 3.50$, $P = 0.034$) and the marginal effect on abundance was 7.56 (standard error = 3.67, $F_{1,48} = 4.24$, $P = 0.022$). The equivalent figures for the *P. fairmairei* manipulation are a log odds of -2.63 (standard error = 1.19, $\chi^2_1 = 7.37$, $P = 0.003$) and a marginal effect on abundance of 2.83 (standard error = 2.4, $F_{1,5} = 1.39$, $P = 0.14$).

Received 24 November 2003; accepted 6 February 2004; doi:10.1038/nature02394.

- Novotny, V. et al. Low host specificity of herbivorous insects in a tropical forest. *Nature* **416**, 841–844 (2002).
- Strong, D. R., Lawton, J. H. & Southwood, T. R. E. *Insects on Plants: Community Patterns and Mechanisms* (Blackwell Science, Oxford, 1984).
- Holt, R. D. Predation, apparent competition and the structure of prey communities. *Theor. Popul. Biol.* **12**, 197–229 (1977).
- Strauss, S. Y. Indirect effects in community ecology—their definition, study and importance. *Trends Ecol. Evol.* **6**, 206–210 (1991).
- Wootton, J. T. The nature and consequences of indirect effects in ecological communities. *Annu. Rev. Ecol. Syst.* **25**, 443–466 (1994).
- Abrams, P. A., Menge, B. A., Mittelbach, G. G., Spiller, D. A. & Yodanis, C. *Food Webs: Integration of Patterns and Dynamics*. (eds Polis, G. & Winemiller, K. O.) 371–395 (Chapman & Hall, New York, 1995).
- Holt, R. D. & Lawton, J. H. Apparent competition and enemy-free space in insect host–parasitoid communities. *Am. Nat.* **142**, 623–645 (1993).
- Schmitt, R. J. Indirect interactions between prey: apparent competition, predator aggregation, and habitat segregation. *Ecology* **68**, 1887–1897 (1987).
- Menge, B. A. Indirect effects in marine rocky intertidal interaction webs: Patterns and importance. *Ecol. Monogr.* **65**, 21–74 (1995).
- Tomkins, D. M., Draycott, R. A. H. & Hudson, P. J. Field evidence for apparent competition mediated via the shared parasites of two gamebird species. *Ecol. Lett.* **3**, 10–14 (2000).
- Muller, C. B. & Godfray, H. C. J. Apparent competition between two aphid species. *J. Anim. Ecol.* **66**, 57–64 (1997).
- Morris, R. J., Müller, C. B. & Godfray, H. C. J. Field experiments testing for apparent competition between primary parasitoids mediated by secondary parasitoids. *J. Anim. Ecol.* **70**, 301–309 (2001).
- Bonsall, M. B. & Hassell, M. P. Apparent competition structures ecological assemblages. *Nature* **338**, 371–373 (1997).
- Settle, W. H. & Wilson, L. T. Invasion by the variegated leafhopper and biotic interactions: parasitism, competition and apparent competition. *Ecology* **71**, 1461–1470 (1990).
- Janzen, D. H. Herbivores and the number of tree species in tropical forests. *Am. Nat.* **104**, 501–528 (1970).
- Hespenheide, H. A. Bionomics of leaf-mining insects. *Annu. Rev. Entomol.* **36**, 535–560 (1991).

17. Lewis, O. T. *et al.* Structure of a diverse tropical forest insect-parasitoid community. *J. Anim. Ecol.* **71**, 855–873 (2002).
18. Holt, R. D. & Lawton, J. H. The ecological consequences of shared natural enemies. *Annu. Rev. Ecol. Syst.* **25**, 495–520 (1994).
19. Abrams, P. The theory of limiting similarity. *Annu. Rev. Ecol. Syst.* **14**, 359–376 (1983).
20. Jeffries, M. J. & Lawton, J. H. Enemy free space and the structure of ecological communities. *Biol. J. Linn. Soc.* **23**, 269–286 (1984).
21. Abrams, P. A. & Matsuda, H. Positive indirect effects between prey species that share predators. *Ecology* **77**, 610–616 (1996).
22. Bonsall, M. B. & Hassell, M. P. The effects of metapopulation structure on indirect interactions in host-parasitoid assemblages. *Proc. R. Soc. Lond. B* **267**, 2207–2212 (2000).
23. Bonsall, M. B. & Hassell, M. P. Parasitoid-mediated effects: apparent competition and the persistence of host-parasitoid assemblages. *Res. Popul. Ecol.* **41**, 59–68 (1999).
24. Turchin, P. *Complex Population Dynamics: A Theoretical/Empirical Synthesis* (Princeton Univ. Press, Princeton, New Jersey, 2003).
25. Power, M. E. *et al.* Challenges in the quest for keystones. *Bioscience* **46**, 609–620 (1996).
26. Howarth, F. G. Environmental impacts of classical biological control. *Annu. Rev. Entomol.* **36**, 485–509 (1991).
27. Simberloff, D. in *Selection Criteria and Ecological Consequences of Importing Natural Enemies* (eds Kauffman, W. C. & Nechols, J. R.) 103–117 (Entomological Society of America, Lanham, Maryland, 1992).
28. Henneman, M. L. & Memmott, J. Infiltration of a Hawaiian community by introduced biological control agents. *Science* **293**, 1314–1316 (2001).
29. Memmott, J. & Godfray, H. C. J. in *Parasitoid Community Ecology* (eds Hawkins, B. A. & Sheehan, W.) 300–318 (Oxford Univ. Press, Oxford, 1994).

Acknowledgements We are particularly grateful to A. Lord. We thank Ch. Bol, Ce. Bol, I. Bol, N. Coc, L. Davies and E. Saquil for support and assistance in the field, and our taxonomic collaborators, including J. LaSalle, D. Quicke and C. Whiteford. Research permission was granted by the Ministry of Natural Resources in Belize, and we are grateful to the Natural History Museum London for allowing us to work at Las Cuevas, and the Natural Environment Research Council for financial support. O.T.L. is a Royal Society University Research Fellow.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.J.M. (r.j.morris@imperial.ac.uk).

Ban on triazine herbicides likely to reduce but not negate relative benefits of GMHT maize cropping

J. N. Perry¹, L. G. Firbank², G. T. Champion³, S. J. Clark¹, M. S. Heard⁴, M. J. May³, C. Hawes⁵, G. R. Squire⁵, P. Rothery⁴, I. P. Woivod¹ & J. D. Pidgeon³

¹Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK

²NERC Centre for Ecology and Hydrology, Lancaster, Cumbria LA1 4AP, UK

³Broom's Barn Research Station, Higham, Bury St Edmunds, Suffolk IP28 6NP, UK

⁴NERC Centre for Ecology and Hydrology, Abbots Ripton, Huntingdon, Cambridgeshire PE28 2LS, UK

⁵Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

The UK Farm-Scale Evaluations (FSE) compared the effects on biodiversity of management of genetically modified herbicide-tolerant (GMHT) spring-sown crops with conventional crop management¹. The FSE reported larger weed abundance under GMHT management for fodder maize², one of three crops studied. Increased seed production may be important for the long-term persistence of these arable weeds and may benefit invertebrates, small mammals and seed-eating birds¹. In three-quarters of FSE maize fields, growers used atrazine on the conventionally managed half, reflecting contemporary commercial practice³. Withdrawal of the triazine herbicides atrazine, simazine and cyanazine from approved lists of EU chemicals⁴ could therefore reduce or even reverse the reported benefits of

GMHT maize^{1,2,5}. Here we analyse effects of applications of triazine herbicides in conventional maize regimes on key indicators⁶, using FSE data. Weed abundances were decreased greatly relative to all other regimes whenever atrazine was applied before weeds emerged. Here, we forecast weed abundances in post-triazine herbicide regimes^{7,8}. We predict weed abundances under future conventional herbicide management to be considerably larger than that for atrazine used before weeds emerged, but still smaller than for the four FSE sites analysed that used only non-triazine herbicides. Our overall conclusion is that the comparative benefits for arable biodiversity of GMHT maize cropping would be reduced, but not eliminated, by the withdrawal of triazines from conventional maize cropping.

Herbicide usage for FSE fodder maize⁹ is shown in Table 1. We studied the response of five key vegetation indicators^{2,5,6} for total weeds, total monocotyledonous weeds (monocots) and total dicotyledonous weeds (dicots) to combinations of herbicide use (Table 2). We found the following consistent trends, exemplified in Fig. 1. Weed abundances were decreased greatly compared to any other regime, whenever atrazine was applied before weeds emerged (pre-emergence⁹, see Methods) on conventional crops. For the other conventional regimes, weed abundance for the few sites that were treated only with non-triazines was consistently slightly greater than those for the triazines. With this exception, weed abundance in fields that were treated conventionally (excluding those where atrazine was used pre-emergence) appeared similar. Weed abundance under GMHT management was usually greater than that for atrazine applied post-emergence or for the other triazines applied pre-emergence.

These trends suggested three independent contrasts, tested formally in Table 3. The first comparison is within the three conventional treatment combinations (in which atrazine was applied post-emergence, AĒ; other triazines were applied pre-emergence, ĀE; or non-triazines alone were applied post-emergence, ĀĒ), that is, excluding those where atrazine was used as a pre-emergence

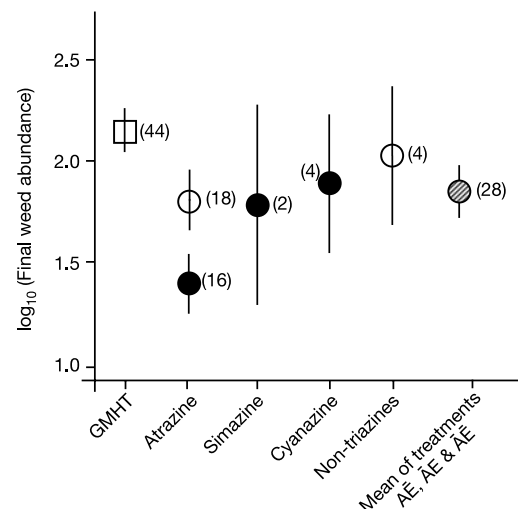


Figure 1 Mean abundance of total pre-harvest weeds and herbicide use. Consistent treatment effects from Table 2, illustrated here by mean abundance of total pre-harvest weeds in FSE fodder-maize per GMHT (square symbol) or conventional (round symbols) half-fields, and treated either with pre-emergence herbicide plus possible post-emergence application(s) (filled symbols, E) or with post-emergence herbicide only (open symbols, Ē). Hatched symbol represents the mean of the three conventional regimes AĒ, ĀE and ĀĒ: that is, all those other than atrazine applied pre-emergence. Numbers in brackets denote *N*, the number of half-fields. Bar represents 95% confidence interval for each mean.

Table 1 **Herbicides applied to 58 FSE fodder-maize conventional half-fields**

	Pre- (plus possibly post-) emergence (E)		Post-emergence only (Ē)		Total
	Number of half-fields	Application rate (g ha ⁻¹) (s.e.m.)	Number of half-fields	Application rate (g ha ⁻¹) (s.e.m.)	Number of half-fields
With atrazine (A)					
A alone	8	1,272 (75)	11	1,336 (62.6)	19
A + non-triazines (NT)	10	1,628 (302)	17	1,244 (97.0)	27
Total atrazine	18	1,470 (172)	28	1,280 (63)	46
Other triazines					
Simazine alone	2	1,150 (0)	0	—	2
Simazine + NT	0	—	0	—	0
Cyanazine alone	0	—	0	—	0
Cyanazine + NT	5	2,233 (163)	0	—	5
Total of other triazines	7	1,924 (229)	0	—	7
NT alone	1	2,115 (—)	4	420 (146)	5
Total of all treatments	26	1,617 (139)	32	1,173 (77.1)	58

Application rates expressed as total grams of active ingredient per hectare from all herbicides combined, averaged over half-fields^a, with standard error of mean in parentheses.

Table 2 **Weed abundances and herbicide use**

Taxa	AE (atrazine, pre-emergence)		AĒ (atrazine, pre-emergence)		ĀĒ (other triazines)		ĀĒ (non-triazines)		GMHT	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Total weeds										
Final plant density	16	1.38	18	1.79	6	1.84	4	2.01	44	2.16
Biomass (g)	14	1.97	16	2.53	5	2.44	4	3.02	39	2.64
Seed rain	18	1.75	25	2.18	6	1.78	4	2.57	53	2.27
Seedbank <i>t</i> + 1	6	1.94	11	2.23	6	2.27	2	2.38	25	2.21
Seedbank change <i>t</i> to <i>t</i> + 1	6	−0.12	11	0.22	6	0.17	2	−0.08	25	0.09
Dicots										
Final plant density	16	1.12	18	1.37	6	1.62	4	1.91	44	1.92
Biomass	14	1.54	16	2.07	5	2.22	4	2.92	39	2.48
Seed rain	18	1.58	25	1.96	6	1.58	4	2.35	53	2.16
Seedbank <i>t</i> + 1	6	1.53	11	1.99	6	2.11	2	2.06	25	1.96
Seedbank change <i>t</i> to <i>t</i> + 1	6	−0.17	11	0.27	6	0.24	2	−0.14	25	0.11
Monocots										
Final plant density	16	0.92	18	1.36	6	1.44	4	1.25	44	1.64
Biomass	14	0.93	16	1.96	5	1.84	4	2.05	39	1.77
Seed rain	16	0.83	19	1.34	6	0.97	4	1.93	45	1.18
Seedbank <i>t</i> + 1	6	1.54	11	1.58	6	1.60	2	2.03	25	1.73
Seedbank change <i>t</i> to <i>t</i> + 1	6	−0.18	11	0.11	6	−0.04	2	−0.05	25	0.03

Weed abundances expressed as means of logarithmically transformed totals of final plant densities, biomass (g), seed rain and seedbank densities, per conventional FSE fodder-maize half-field sample, treated either with (A) or without (Ā) atrazine, under regimes that included either pre-emergence (E) herbicide or purely post-emergence (Ē) application, and in GMHT half-fields. *t* relates to year crop was in ground; *t* + 1 relates to subsequent year.

Table 3 **Tests of three independent contrasts of the mean weed abundances described in Fig. 1 and Table 2**

Taxa	Within treatments AE, ĀE & ĀĒ (2 d.f.)	AE versus rest (1 d.f.)			GMHT versus mean of AE, ĀE & ĀĒ (1 d.f.)				
	<i>P</i>	<i>d</i>	s.e.(<i>d</i>)	<i>P</i>	<i>d</i>	s.e.(<i>d</i>)	<i>P</i>	<i>R</i> ₀	<i>R</i>
Total weeds									
Final plant density	0.199	−0.65	0.080	<0.001	0.33	0.070	<0.001	3.08	2.16
Biomass	0.479	−0.65	0.161	0.001	0.05	0.139	0.916	1.82	1.12
Seed rain	0.915	−0.48	0.202	0.024	0.11	0.170	0.720	1.87	1.29
Seedbank <i>t</i> + 1	0.936	−0.29	0.091	0.474	−0.05	0.064	0.860	1.07	0.90
Seedbank change <i>t</i> to <i>t</i> + 1	0.976	−0.24	0.103	0.433	−0.09	0.072	0.478	0.96	0.82
Dicots									
Final plant density	0.180	−0.64	0.099	<0.001	0.42	0.086	<0.001	3.52	2.61
Biomass	0.397	−0.84	0.217	<0.001	0.24	0.189	0.384	3.02	1.76
Seed rain	0.835	−0.49	0.216	0.012	0.23	0.181	0.456	2.32	1.68
Seedbank <i>t</i> + 1	0.616	−0.46	0.097	0.032	−0.08	0.068	0.598	1.1	0.83
Seedbank change <i>t</i> to <i>t</i> + 1	0.730	−0.32	0.103	0.078	−0.11	0.072	0.218	0.97	0.79
Monocots									
Final plant density	0.251	−0.61	0.094	<0.001	0.28	0.082	0.007	2.88	1.91
Biomass	0.657	−0.91	0.206	0.016	−0.19	0.179	0.642	1.58	0.65
Seed rain	0.790	−0.41	0.179	0.124	−0.17	0.154	0.317	1.18	0.68
Seedbank <i>t</i> + 1	0.217	−0.15	0.146	0.773	0.09	0.102	0.132	1.31	1.24
Seedbank change <i>t</i> to <i>t</i> + 1	0.648	−0.22	0.135	0.743	−0.02	0.095	0.775	1.09	0.96

Each test has a *P*-value. For the last two contrasts, the mean difference, *d*, and standard error of logarithmically transformed abundance is given. For the third contrast the value of *R* = 10² is also calculated, for comparison with the original GMHT versus conventional treatment-effect ratio, *R*₀, repeated from ref. 2. d.f., degrees of freedom. s.e.(*d*), standard error of the difference, *d*.

herbicide (AE). The second comparison is between AE and all the other treatment combinations (AĒ, ĀE, ĀĒ and GMHT). The third comparison, which is independent of the first two, is between GMHT and treatments AĒ, ĀE and ĀĒ.

Weed abundance for atrazine applied pre-emergence in conventional crops was usually statistically significantly less than the mean of all other regimes (Table 3). Typically, it was reduced by two-thirds; monocot biomass was reduced eightfold. Differences in weed abundance never approached significance for the tests among the three other conventional herbicide regimes (AĒ, ĀE, ĀĒ).

The treatment effect between GMHT and conventional crops in the FSE was presented^{1–3,5} as the multiplicative ratio of the geometric-mean half-field GMHT abundance divided by that for the conventional. The value of this multiplicative ratio, R , as defined here, contrasts GMHT with the conventional regimes, excluding atrazine applied pre-emergence. The value previously reported in ref. 2, termed R_0 here, applies to all data including sites with pre-emergence atrazine. Regression analysis of results in Table 3 confirmed that R was a consistent proportion: approximately 0.687 (with s.e. 0.026) of R_0 .

These results were consistent with the effects reported previously² for GMHT management in maize, which produces a larger number of smaller plants, typically half as large in the reproductive phase, relative to conventional management. All of the more abundant species of weeds in the FSE have seeds that persist in seedbanks for several seasons⁵, so the effects of weed management in one season are buffered¹. Our results for seedbanks have relatively small sample size (N , Table 2), but appear to confirm this buffering effect.

Caution is required in predicting the magnitude of effects on biodiversity of future herbicide regimes⁷, but considerable information exists on the probable replacements for triazines in conventional maize in the UK⁹ to aid forecasts. The larger weed abundances for GMHT than for conventional maize management reported previously², and expressed through the multiplicative treatment-effect ratio (R_0), were caused partly by the effectiveness of atrazine used pre-emergence (AE). Following withdrawal of triazines in the EU, other conventional herbicides will almost certainly be less effective than atrazine used pre-emergence and values of R for future herbicide management for maize crops would therefore be reduced^{1,8}. Future conventional management of UK maize crops will probably involve⁸: (1) a greater range of herbicides than in the FSE (Table 1; ref. 9); (2) more use of sulphonyl urea herbicides, together with pendimethalin, bromoxynil and fluroxypyr; (3) non-triazine herbicides with greater residual activity; and (4) non-triazine herbicides applied (unlike the FSE data reported here) pre-emergence.

We therefore expect weed abundances under future conventional herbicide management for maize crops to be considerably larger than that reported here for atrazine used pre-emergence (AE), but smaller than for the four sites analysed that used non-triazines alone. A reasonable but necessarily tentative approximation to future weed abundances is therefore the mean of the pooled categories (AĒ, ĀE, ĀĒ): that is, all sites excluding those where atrazine was used pre-emergence.

If this pooled category of herbicide regimes is indeed representative of weed control in post-triazine conventional crops, and if the weed management in GMHT maize remains the same as observed within the FSE, then final weed numbers would still be larger in GMHT than in conventional maize. Dicot weed biomass would have a value of $R = 1.76$, well above unity, although not significantly so; by contrast, monocot weed biomass would be smaller under GMHT than under future conventional management. The predicted effects on seed rain (amount of seed shed by a plant) would largely follow those for biomass, with more than 1.5 times as many expected under GMHT management, but with a relatively large standard error. Generally, the statistical significance of predicted differences for individual indicators is of less importance than the

biological significance of the consistency^{1–3} of the relation between R and R_0 over all indicators.

To place these results in context, we reiterate¹ that reported differences in the FSE were as large between crops as between treatments. Conventional maize was the poorest combination for biodiversity, and even under GMHT management seed rain only marginally exceeded that for conventional beet and GMHT spring oilseed rape². However, we conclude that the comparative benefits for arable biodiversity of GMHT maize cropping would be reduced, but not eliminated, by the withdrawal of triazines from conventional maize cropping. □

Methods

Table 1

Of 58 FSE fodder-maize conventional half-fields⁹, 46 received atrazine (A), with or without other non-triazines; 12 received no atrazine (Ā). The 26 conventional crops that were treated pre-emergence (defined as treated within 14 days after sowing for crops sown on or before 15 May, or within 7 days for crops sown after 15 May) are distinguished from the 32 with purely post-emergence (Ē) application. Pre-emergence herbicide use almost always included a triazine, usually atrazine, and was often followed by additional post-emergence herbicide application(s). Besides atrazine, other triazines used were simazine on two half-fields and cyanazine on five. Five half-fields received a non-triazine (NT) alone. Application rates are expressed as total amount of active ingredient (g ha^{-1}) from all herbicides combined, averaged over half-fields⁹, with standard error of mean. When one triazine herbicide was used, no other was; however, other non-triazine herbicides were usually used in addition to a triazine. (The use of a non-triazine in addition to the use of a triazine appeared to have little effect on weed abundance and is not considered further here.) Measurement of the seedbank was done in the top 15 cm of soil, using samples usually taken before any herbicide was applied⁷. There was no significant dependence of the use of either atrazine or of a pre-emergence herbicide on the size of the initial seedbank. For crops treated with atrazine, the amount of active ingredient applied⁹ (g ha^{-1}) was very similar between the pre-emergence (E) and post-emergence only (Ē) treatments, whether atrazine was used in combination with a non-triazine or not.

Table 2

We studied the response of five key vegetation indicators^{2,5} for total weeds, total monocots and total dicots to combinations of herbicide use. Values for seedbank change are transition rates per half-field. All indicators and calculations are as described in ref. 2, where weed abundances are expressed as means of logarithmically transformed abundances per FSE half-field sample. Sample size, N , may be reduced from the values in Table 1 because data were missing for a half-field or abundance was too small for analysis^{2,3,9}.

Figure 1

We found consistent trends in Table 2, exemplified in Fig. 1. Fourteen fields either had missing data or abundance was too small for analysis. Of the 44 remaining conventional half-fields, 34 received atrazine with or without other non-triazines (A), and ten received no atrazine (Ā). Of these ten, two received simazine alone, four received cyanazine and another non-triazine, and four received non-triazines alone. Confidence intervals were based on residual variation after removal of site and treatment effects.

Table 3

The mean difference between logarithmically transformed abundances for each contrast examined is termed d . Analyses were regressions where treatment effects were estimated after removal of site effects. Standard errors of d were therefore based on residual variation after removal of both site and treatment effects. The treatment effect, R , for contrasts involving GMHT versus conventional regimes, is presented as the multiplicative ratio of the half-field GMHT abundance divided by that for the conventional, calculated from $R = 10^d$. For each of the three independent contrasts shown, the P -value, from analysis of variance³, is based on an F -test. The third contrast, intended to compare GMHT with possible future management, is between GMHT and treatments AĒ, ĀE and ĀĒ; for this the value of $R = 10^d$ is also calculated, for comparison with the original GMHT versus conventional treatment-effect ratio³, R_0 , repeated from ref. 2. This contrast was usually based on $N > 25$ sites. Data yet to be analysed^{1,2,5} concerning seedbank densities during the first and second year after the treatments, are still required to establish reliable rates of change of maize seedbanks over full rotations.

Received 12 January; accepted 27 January 2004; doi:10.1038/nature02374.

Published online 5 March 2004.

1. Firbank, L. G. *et al.* The Implications of Spring-Sown Genetically Modified Herbicide-Tolerant Crops for Farmland Biodiversity: A Commentary on 'The Farm Scale Evaluations Of Spring Sown Crops'. (Defra, 2003); at (<http://www.defra.gov.uk/environment/gm/fse/index.htm>).
2. Heard, M. S. *et al.* Weeds in fields with contrasting conventional and genetically modified herbicide-tolerant crops. 1. Effects on abundance and diversity. *Phil. Trans. R. Soc. Lond. B* **358**, 1819–1832 (2003).
3. Perry, J. N., Rothery, P., Clark, S. J., Heard, M. S. & Hawes, C. Design, analysis and power of the Farm-Scale Evaluations of Genetically-Modified Herbicide-Tolerant crops. *J. Appl. Ecol.* **40**, 17–31 (2003).
4. Pesticide Safety Directorate. *EC Review Programme for Existing Active Substances*. (PSD, York, 2003);

at (http://www.pesticides.gov.uk/ec_process/ECreviews/EC_review_programme.htm).

5. Heard, M. S. *et al.* Weeds in fields with contrasting conventional and genetically modified herbicide-tolerant crops. 2. The effects on individual species. *Phil. Trans. R. Soc. Lond. B* **358**, 1833–1846 (2003).
6. Firbank, L. G. *et al.* Farm-scale evaluation of genetically modified crops. *Nature* **399**, 727–728 (1999).
7. British Statutory Nature Conservation Agencies. *Advice to ACRE on the Implications of the Farm Scale Evaluations for Biodiversity in the UK*. Evidence submitted to the ACRE Committee's Farm-scale Evaluation Results Open Meetings, November 2003 (English Nature, Peterborough, 2003); at (<http://www.livegroup.co.uk/acrefarmscaleevaluations/SSL/index2.php?page=submissions>).
8. Marshall, J. *Glufosinate-tolerant Maize: Implications of the USA Experience for Weed Control in Forage Maize in the UK*. Appendix 2 of evidence submitted by Greenpeace to the ACRE Committee's Farm-scale Evaluation Results Open Meetings, November 2003 (Greenpeace UK, London, 2003); at (<http://www.livegroup.co.uk/acrefarmscaleevaluations/SSL/index2.php?page=submissions>).
9. Champion, G. T. *et al.* Crop management and agronomic context of the Farm Scale Evaluations of genetically modified herbicide-tolerant crops. *Phil. Trans. R. Soc. Lond. B* **358**, 1801–1818 (2003).

Acknowledgements We thank members of the Scientific Steering Committee of the FSE for their support. A. Tuse provided suggestions. The FSE were funded by Defra and the Scottish Executive

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to J.N.P. (joe.perry@bbsrc.ac.uk).

Cortical activity reductions during repetition priming can result from rapid response learning

Ian G. Dobbins¹, David M. Schnyer^{2,3}, Mieke Verfaellie² & Daniel L. Schacter⁴

¹Psychological and Brain Sciences, Duke University, Durham, North Carolina 27708, USA

²Memory Disorders Research Center, Boston VA Healthcare System and Boston University School of Medicine, Boston, Massachusetts 02118, USA

³Athinoula A. Martinos Center for Biomedical Imaging, MGH/MIT/HMS, Charlestown, Massachusetts 02129, USA

⁴Department of Psychology, Harvard University, Cambridge, Massachusetts 02138, USA

Recent observation of objects speeds up their subsequent identification and classification^{1,2}. This common form of learning, known as repetition priming, can operate in the absence of explicit memory for earlier experiences^{3,4}, and functional neuroimaging has shown that object classification improved in this way is accompanied by 'neural priming' (reduced neural activity) in prefrontal, fusiform and other cortical regions^{5–10}. These observations have led to suggestions that cortical representations of items undergo 'tuning', whereby neurons encoding irrelevant information respond less as a given object is observed repeatedly¹⁰, thereby facilitating future availability of pertinent object knowledge. Here we provide experimental support for an alternative hypothesis, in which reduced cortical activity occurs because subjects rapidly learn their previous responses¹¹. After a primed object classification (such as 'bigger than a shoebox'), cue reversal ('smaller than a shoebox') greatly slowed performance and completely eliminated neural priming in fusiform cortex, which suggests that these cortical item representations were no more available for primed objects than they were for new objects. In contrast, prefrontal cortex activity tracked behavioural priming and predicted the degree to which cue reversal would slow down object classification—highlighting the role of the prefrontal cortex in executive control.

We contrasted tuning and response-learning hypotheses by examining the response sensitivity to a cue change following primed object-size judgements, under the assumption that a response-

learning strategy would be abandoned when prior responses became inappropriate, even if retrieval remained directed towards the same object properties¹² (Fig. 1). Importantly, the cue change did not alter the target knowledge (object size) or the comparison item (the shoebox); it merely rendered the previous responses inappropriate. If information regarding target object properties becomes increasingly available with priming (that is, tuning occurs), then this cue change should have little effect on performance (speed of object size classification) or on neural priming, because the same object knowledge is targeted by either cue. In contrast, if subjects rapidly come to rely upon learned response associations, then the switch should generate considerably slowed performance in response time and re-engagement of the same cortical networks employed during initial processing should occur.

Behaviourally, reaction times (RT) were examined for cue type (start, switch, return) and priming type (novel, low, high) within

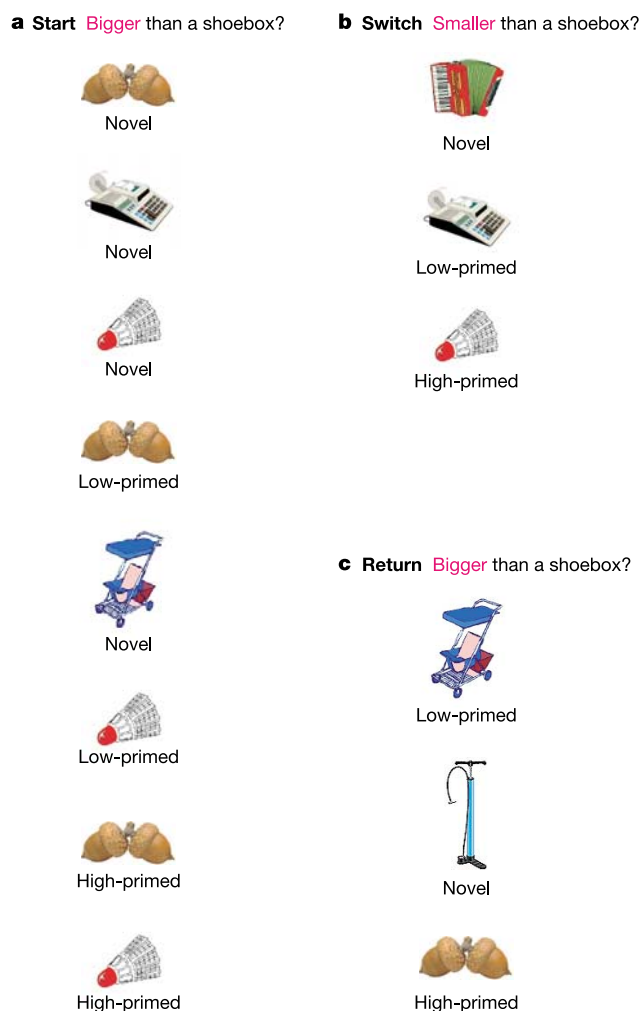


Figure 1 Experiment design. Subjects encountered three sequential phases during scanning ('Start', 'Switch' and 'Return'), during which the retrieval cue direction ('bigger than' or 'smaller than' judgement, needing a yes/no response) was reversed or consistent with respect to the initial start phase cue. **a**, During the start phase, objects were seen either once ('Novel'), or three times ('High-primed'). **b**, During the subsequent switch phase, half of the items from the previous start phase that had initially been classified as novel or high-primed were re-presented along with a new set of novel items. Thus the switch phase contained items that were viewed for the first time (novel), items viewed for the second time ('Low-primed') and items making their fourth appearance (high-primed). **c**, Similarly, the final return phase used the remaining novel and high-primed items from the initial start phase along with a completely new set of novel items.

each cue type ($F_{2,4,35.9} = 5.75$, $P < 0.005$) (Fig. 2a). Although RT to novel stimuli did not vary with cue type ($F_{1,4,21.2} = 2.7$, $P > 0.10$), low- and high-primed RT values slowed during the switch, and partially recovered when the cue was returned to the original form for the remaining items (return phase), as confirmed by quadratic trends for high-primed ($F_{1,15} = 31.50$, $P < 0.0001$) and low-primed ($F_{1,15} = 16.52$, $P < 0.01$) items across cue type. Although noticeably slowed by the cue switch, priming was not completely eliminated (high- and low-primed RTs versus novel, P values < 0.001).

Event-related functional magnetic resonance imaging (fMRI) demonstrated reduced cortical activity in prefrontal, parietal, occipito-temporal and fusiform regions for (both low- and high-) primed relative to novel items during the start phase, and an interaction demonstrating significant disruption of this neural priming for the high-primed items in prefrontal cortex (PFC), fusiform and other regions as a result of the cue reversal. Priming disruptions were more evident in the left hemisphere, and potentially correspond to the behavioural RT changes (Fig. 3).

To determine whether regional activity paralleled response behaviour, regions of interest (ROIs) were extracted from regions near those implicated in previous priming studies^{7,13} that demonstrated a start-switch phase interaction, namely, left posterior PFC (BA 9/44) and left fusiform (BA 37). For these regions, the mean priming signals (novel minus high- or low-primed) were subjected to quadratic trend analysis (Fig. 3c). For posterior PFC, the trend was marginally significant for the high-primed condition ($F_{1,15} = 3.98$, $P = 0.065$), suggesting a significant decline in the priming signal during the cue switch that only partially recovered during the return phase. The priming signal for the low-primed condition in PFC, although demonstrating a significant decline during the switch phase, showed no sign of recovery during the return phase (quadratic trend $P > 0.29$). The trends for the fusiform region priming responses were more robust (high-primed $F_{1,15} = 17.74$, $P < 0.001$; low-primed $F_{1,16} = 4.61$, $P < 0.05$), suggesting elimination of neural priming followed by recovery of the neural difference between novel and primed items when the cue was restored to the original form for both high- and low-primed items respectively. This pattern rules out simple causes such as forgetting and shows that priming-like reductions in fusiform activity were heavily dependent on the stability of the associated response, and not on the requirement to access the same item knowledge across exposures¹⁴.

Although PFC and fusiform reductions are common in priming

studies, they have not been reliably linked to individual performance. To investigate this issue, the mean priming signals (novel minus high-primed) during the start phase were entered into a multiple regression to examine their relation to each subject's behavioural priming scores for the same items (Table 1). Both regions uniquely predicted priming scores, accounting for 63% of the variance. Additionally, the same signal differences were also used in an attempt to predict the subsequent behavioural cost of switching the cue. Only the posterior PFC region predicted switching costs, with regression accounting for 59% of the variance (Table 1).

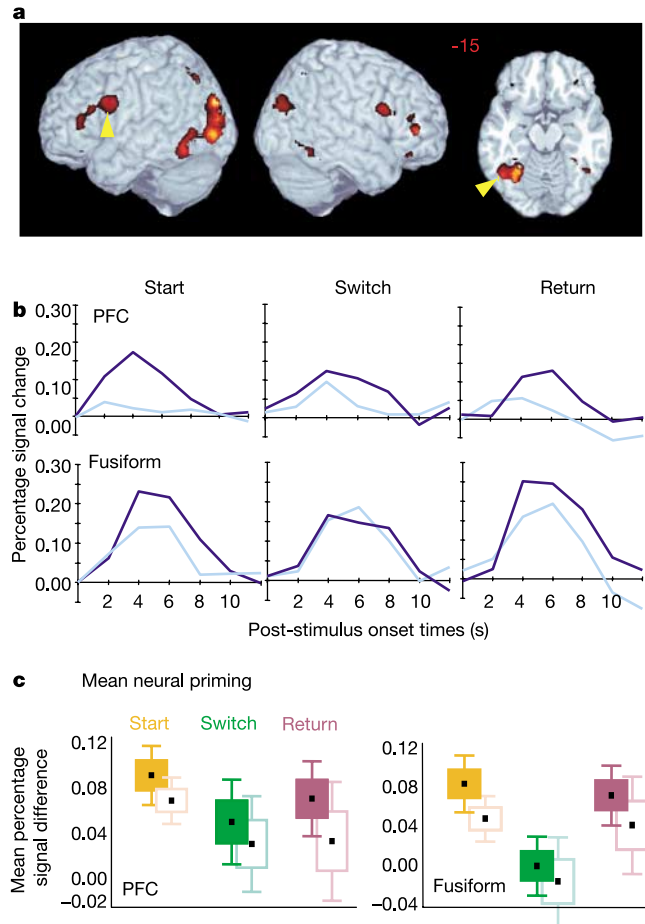


Figure 3 SPM results for novel versus high-primed responses across start, switch and return phases. During the start phase, neural priming was evident in large regions of the PFC, parietal, inferotemporal and fusiform areas. A minor change in cue direction ('bigger than' to 'smaller than') during the switch phase disrupted neural priming, particularly in posterior regions. Returning the cue to its original form (return phase) resulted in partial recovery of neural priming, especially noticeable in the fusiform and other posterior regions. **a**, SPM interaction map showing regions that demonstrated a significant disruption of neural priming signal in the switch relative to start phase overlaid on a high-resolution three-dimensional canonical brain in Montreal Neurological Institute (MNI)152 space and displayed on an axial slice (-15), both thresholded at 0.001. Yellow arrows denote approximate location of ROIs used for the time courses and box plots. **b**, Haemodynamic time courses. Dark blue lines indicate response to novel items; pale blue lines indicate response to high-primed items. The left PFC region consisted of 80 voxels (MNI -45, 6, 27; BA 9/44). The left fusiform region consisted of 62 voxels (MNI -24, -57, -15; BA 19/37). **c**, Box plots indicate mean percentage signal difference between novel and high-primed (filled boxes) and novel and low-primed (open boxes) items for each phase, averaged from 4 to 8 s post-stimulus onset for start, switch and return phases. Box indicates one standard error of the between-subjects mean; box plus error bars indicates two standard errors.

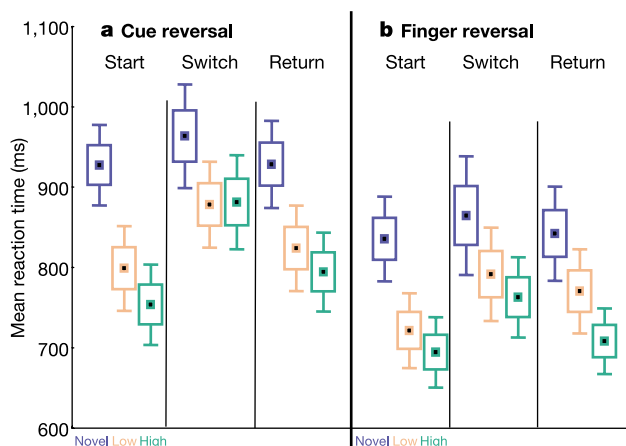


Figure 2 Behavioural reaction time data for scanned and finger-reversal experiments as a function of cue phase. Box indicates one standard error of the between-subjects mean; box plus error bars indicates two standard errors.

Thus in posterior PFC, a greater neural response reduction indicates larger behavioural facilitation, and anticipates a larger relative impairment when subjects encounter a cue that still targets the same semantic content, but is modified such that the previous response has become inappropriate. This observation supports the contention that reductions in this region do not indicate increasingly available size information, since this would not anticipate larger costs when switching the cue. Instead, the PFC data indicate a shift from controlled working memory/executive operations thought to be mediated by this area¹⁵, to the reliance on more automatic learned response associations.

In contrast to PFC, the activity difference in left fusiform yielded a negative relation with priming scores in the start phase, and was unrelated to the later cue-reversal performance cost. This inverse relation to priming scores probably resulted from the negative relation of fusiform activity to RTs for novel items across subjects (fusiform $r = -0.64$, $P < 0.01$) with increased fusiform activity associated with quicker, not slower, responding. This negative relation was not due to general cortical arousal differences across subjects because it remained even when the left PFC region was also added to the regression, suggesting that increased fusiform activity predicted rapid novel-item RTs (Table 1). By the third item repetition however, fusiform activity no longer inversely predicted RT. This suggests that the negative regression weight of the fusiform when predicting priming scores occurred primarily because of the initial association between greater fusiform activity and quicker responding to novel items, a pattern no longer present by the third trial, by which time responses were presumably well learned (Table 1).

This negative relation between fusiform response and subject responses to novel stimuli has not previously been reported and its meaning with respect to the role of fusiform in long-term object-knowledge is unclear. An informal item analysis suggested that, for each subject, RTs to novel stimuli were largely governed by the size difference between the referent (the shoebox) and probe, with distant probes (such as a pencil or strawberry) responded to more quickly than close ones (such as Kleenex box or dustpan). However, fusiform activity did not parametrically change with RTs to individual novel stimuli, even at liberal thresholds, suggesting that it was insensitive to the trial-by-trial difficulty of the judgement.

Another possibility is that the mean fusiform response for each subject was governed by the overall availability of pre-existing object representations¹⁶. If so, then subjects with more experience of many

of the items would be expected to show more activity and respond more quickly. Although speculative, this explanation fits with evidence that similar fusiform regions increase response during the gradual acquisition of expertise with objects¹⁷. In contrast to the null parametric fusiform response to trial difficulty, posterior PFC demonstrated a robust positive parametric relation with individual novel-item RTs—this again suggests a trial-by-trial role for PFC in executive mechanisms that are increasingly critical as referent and probe sizes approach one another.

Overall, the data strongly suggest that subjects rapidly learn their responses. To examine the specificity of this learning, we examined whether similar slowing would occur when just the finger assignment for 'yes' and 'no' responses was reversed, but cue direction remained constant ('bigger than a shoebox?') using a non-scanned group of participants ($n = 16$). A similar interaction between cue type and priming type was observed ($F_{2,7, 40.9} = 3.91$, $Mse = 991.35$, $P < 0.05$). The finger-reversal manipulation did not affect novel-item RTs ($F_{2,0, 29.8} = 1.97$, $P > 0.15$), but did yield a quadratic trend for high- and low-primed item RTs across cue type ($F_{1,15} = 53.18$, $P < 0.0001$; $F_{1,15} = 12.68$, $P < 0.01$) (Fig. 2b). On average, subjects were quicker in the finger-reversal task than in the scanner ($F_{1,30} = 5.72$, $P < 0.05$); however, there were no interactions across the groups. These findings suggest that response learning extends to the level of finger assignment and are consistent with previous behavioural observations of extreme specificity ('hyperspecificity') in the expression of some repetition priming effects^{18,19}.

Our findings demonstrate that a considerable proportion of the behavioural and cortical consequences of object-knowledge priming paradigms may not result from increasing availability of object-related knowledge, as suggested by tuning explanations of neural priming^{10,20}. Instead we suggest a different conceptualization for repetition-related neural-activity reductions¹¹. Using response learning to explain neural priming, we suggest that subjects bypass recovery and analysis of detailed object knowledge whenever the task can be more easily accomplished via a learned association between object identity and prior response. This association was well established after three exposures (high-primed), but the response of the fusiform region, and the behavioural data, suggest that some learning may have occurred even after only a single exposure (low-primed), although the effect was clearly less robust.

The response-learning explanation we offer is probably not directly applicable to previous designs that have not required classification¹⁶, or that have had orthogonal response requirements across encounters^{7,21–23}. However, more generally, signal reductions across a range of priming tasks may be due to a shift in strategy that results in rapidly bypassing local neuronal systems rather than tuning them on the basis of short-term environmental contingencies.

The response-learning mechanism outlined here would conserve capacity-limited executive functions by eliminating the need to recover and evaluate detailed representations of object-related knowledge, and would make the practised version of the task fundamentally different from the initial task⁵. The ability to rapidly learn one's prior responses to recurrent stimuli has strong adaptive value and may be a largely unavoidable consequence of deliberative processing in the service of response goals, particularly when the response set is limited and therefore easily learned, and the items are relatively distinctive. Given this perspective, evidence for rapid changes in the representation of object knowledge as a function of task repetition will require the study of subjects who are unable to learn responses effectively, or experimental designs in which such a mechanism is implausible. Only under such conditions can we be confident that activity reductions in fusiform or related posterior structures are not the indirect result of the response-learning mechanism demonstrated here. □

Table 1 Regression results for predicting behavioural scores with cortical responses

Dependent variable	Region	Beta	Error	t-test value for beta weight	P value	R ²
Priming score*	Left PFC	0.75	0.174	4.33	0.0008	0.63
	Left fusiform	-0.52	0.174	-3.01	0.0099	
Switch cost†	Left PFC	0.79	0.184	4.27	0.0009	0.59
	Left fusiform	-0.29	0.184	-1.57	0.1412	
Novel RT‡	Left PFC	-0.14	0.219	-0.41	0.6822	0.41
	Left fusiform	-0.61	0.219	-2.80	0.0153	
High-prime RT§	Left PFC	-0.21	0.25	-0.83	0.4227	0.21
	Left fusiform	-0.37	0.25	-1.46	0.1681	

For each subject, independent variables in the regressions were constructed using the response of left PFC and fusiform regions from 4–8 s post-stimulus onset. These were then used to predict different behavioural scores in four separate regression analyses as outlined below. Boldface text denotes variables that were significant in each of the four regressions.

*Independent variables were the mean response difference for novel and high-primed items during the start phase. The dependent variable was the corresponding behavioural priming score: novel RT minus high-primed RT during the start phase.

†Independent variables were the same as above. The dependent variable was the switch cost, defined as the difference in behavioural priming scores between switch and start phases: start (novel RT minus high-primed RT) minus switch (novel RT minus high-primed RT).

‡Independent variables were the mean response for novel stimuli relative to baseline during the start phase. The dependent variable was RT to novel stimuli during the start phase.

§Independent variables were the mean response to high-primed stimuli relative to baseline during the start phase. The dependent variable was RT to high-primed stimuli during the start phase.

Methods

Subjects

Sixteen right-handed, 18- to 35-year-old, native-English-speaking volunteers were paid participants (\$50). Informed consent was obtained, as required by the Human Subjects Research Committee at Massachusetts General Hospital.

Materials

Materials consisted of 408 colour drawings of common objects used in previous studies of object-size priming²⁴.

Study procedure

The overall design employed four scans with three experimental phases within each scan. The initial start phase was designed to closely match a standard object-priming paradigm and yield the expected neural and behavioural priming effects.

During the start phase subjects indicated via key-press whether each displayed picture was bigger than an average-sized shoebox ('bigger than a shoebox?') using the left key for a yes response and the right key for a no response. Items presented during this phase were seen once (novel) or three times (high-prime).

Subsequently, for the switch phase, and during the same scanning run, this cue was replaced with a 'smaller than a shoebox?' cue requiring subjects to respond to novel items, and half of the novel and high-primed items from the previous phase. This critical phase was predicted either to incur a cost if subjects were using response learning, or to have very little effect if priming was the result of increasingly efficient size knowledge retrieval (that is, tuning). Behaviourally, subjects reversed 92 and 95% of their responses in the switch phase for low- and high-primed items respectively.

Finally, for the return phase, during the same scanning run, the cue was returned to the original ('bigger than a shoebox?'), and subjects again responded to novel items, and the remaining novel and high-primed items from the initial phase. The behavioural data were Greenhouse-Geisser-corrected for sphericity violations. During each trial, the item remained on the screen for 2,200 ms. The experimental trials were interspersed with blank trials, using an optimal sequencing program²⁵. Each scan lasted 10.5 min.

Data acquisition

Scanning was performed on a 3T Siemens Trio system using a standard head coil. Functional data were collected using a gradient-echo echo-planar pulse sequence (TR = 2,200, TE = 30 ms, 31 axial slices parallel to the AC-PC plane, 3.125 × 3.125 mm, 0.3 mm interslice gap). Before functional data collection, four dummy volumes were discarded to allow for T1 equilibration. Head motion was restricted using a pillow and foam inserts.

fMRI data analysis

Data were preprocessed using statistical parametric mapping—SPM99 (Wellcome Department of Cognitive Neurology, London). Slice-acquisition timing was corrected by resampling slices in time relative to the middle slice collected, followed by a rigid-body motion correction across all runs. Functional data were spatially normalized to a canonical EPI template using a 12-parameter affine and nonlinear cosine basis function transform. Volumes were then spatially smoothed with an 8-mm full-width at half-maximum (FWHM) gaussian kernel. Each session was rescaled such that the mean signal was 100.

The data were analysed by treating subjects as a random effect. Volumes were treated as a temporally correlated time series and modelled using a canonical haemodynamic response function and its first-order time derivative for each event onset. These functions were used along with a basis set of cosine functions that were used to high-pass filter the data, and a covariate representing session effects. The least squares parameter estimates of the response for each condition were then used to construct contrast images based on the canonical HRF for each subject. These images were tested against a null of no difference between contrast conditions. Regions were considered significant, and potentially submitted to a later ROI analysis, if they exceed five or more contiguous voxels at an alpha threshold of 0.001.

ROIs were extracted using peristimulus averaging for voxels significant in the canonical model within an 8-mm-radius sphere of SPM-identified maxima selected based on prior literature.

Received 8 August 2003; accepted 4 February 2004; doi:10.1038/nature02400.

Published online 29 February 2004.

1. Tulving, E. & Schacter, D. L. Priming and human memory systems. *Science* **247**, 301–306 (1990).
2. Toth, J. P. & Reingold, E. M. in *Implicit Cognition* (ed. Underwood, G. D. M.) 41–84 (Oxford Univ. Press, London, 1996).
3. Squire, L. R. & McKee, R. Influence of prior events on cognitive judgments in amnesia. *J. Exp. Psychol.* **18**, 106–115 (1992).
4. Schacter, D. L., Chiu, C. Y. P. & Ochsner, K. N. Implicit memory: A selective review. *Annu. Rev. Neurosci.* **16**, 159–182 (1993).
5. Raichle, M. E. et al. Practice-related changes in human brain functional anatomy during nonmotor learning. *Cereb. Cortex* **4**, 8–26 (1994).
6. Demb, J. B. et al. Semantic encoding and retrieval in the left inferior prefrontal cortex: a functional MRI study of task difficulty and process specificity. *J. Neurosci.* **15**, 5870–5878 (1995).
7. Buckner, R. L. et al. Functional anatomical studies of explicit and implicit memory retrieval tasks. *J. Neurosci.* **15**, 12–29 (1995).
8. Wagner, A. D., Desmond, J. E., Demb, J. B., Glover, G. H. & Gabrieli, J. D. E. Semantic repetition priming for verbal and pictorial knowledge: A functional MRI study of left inferior prefrontal cortex. *J. Cogn. Neurosci.* **9**, 714–726 (1997).
9. Schacter, D. L. & Buckner, R. L. Priming and the brain. *Neuron* **20**, 185–195 (1998).
10. Wiggs, C. L. & Martin, A. Properties and mechanisms of perceptual priming. *Curr. Opin. Neurobiol.* **8**, 227–233 (1998).

11. Logan, G. D. Repetition priming and automaticity: Common underlying mechanisms? *Cognit. Psychol.* **22**, 1–35 (1990).
12. Vriezen, E. R., Moscovitch, M. & Bellos, S. A. Priming effects in semantic classification tasks. *J. Exp. Psychol.* **21**, 933–946 (1995).
13. Koutstaal, W. et al. Perceptual specificity in visual object priming: functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia* **39**, 184–199 (2001).
14. Thompson-Schill, S. L., D'Esposito, M. & Kan, I. P. Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron* **23**, 513–522 (1999).
15. Kane, M. J. & Engle, R. W. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonom. Bull. Rev.* **9**, 637–671 (2002).
16. Henson, R., Shallice, T. & Dolan, R. Neuroimaging evidence for dissociable forms of repetition priming. *Science* **287**, 1269–1272 (2000).
17. Gauthier, I. & Nelson, C. A. The development of face expertise. *Curr. Opin. Neurobiol.* **11**, 219–224 (2001).
18. Schacter, D. L. in *Memory Systems of the Brain* (eds Weinberger, N. M., McGaugh, J. L. & Lynch, G.) 351–379 (The Guilford Press, New York, 1985).
19. Hayman, C. G. & Tulving, E. Is priming in fragment completion based on a "traceless" memory system? *J. Exp. Psychol.* **15**, 941–956 (1989).
20. Henson, R. N. Neuroimaging studies of priming. *Prog. Neurobiol.* **70**, 53–81 (2003).
21. Squire, L. R. et al. Activation of the hippocampus in normal humans: a functional anatomical study of memory. *Proc. Natl Acad. Sci. USA* **89**, 1837–1841 (1992).
22. Schacter, D. L. et al. Conscious recollection and the human hippocampal formation: evidence from positron emission tomography. *Proc. Natl Acad. Sci. USA* **93**, 321–325 (1996).
23. Henson, R. N. et al. Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cereb. Cortex* **13**, 793–805 (2003).
24. Simons, J. S., Koutstaal, W., Prince, S., Wagner, A. & Schacter, D. Neural mechanisms of visual object priming: Evidence for perceptual and semantic distinctions in fusiform cortex. *Neuroimage* **19**, 613–626 (2003).
25. Dale, A. M. Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* **8**, 109–114 (1999).

Acknowledgements We thank L. Nicholls for help with data collection and analysis. This research was supported by grants from the NIMH (D.L.S. and D.M.S.), NINDS (M.V.) and the NIA (D.L.S.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.G.D. (ian.dobbins@duke.edu).

Protein-only transmission of three yeast prion strains

Chih-Yen King & Ruben Diaz-Avalos

Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, USA

Key questions regarding the molecular nature of prions are how different prion strains can be propagated by the same protein and whether they are only protein^{1–3}. Here we demonstrate the protein-only nature of prion strains in a yeast model, the [PSI] genetic element that enhances the read-through of nonsense mutations in the yeast *Saccharomyces cerevisiae*^{4,5}. Infectious fibrous aggregates containing a Sup35 prion-determining amino-terminal fragment labelled with green fluorescent protein were purified from yeast harbouring distinctive prion strains. Using the infectious aggregates as 'seeds', elongated fibres were generated *in vitro* from the bacterially expressed labelled prion protein. *De novo* generation of strain-specific [PSI] infectivity was demonstrated by introducing sheared fibres into uninfected yeast hosts. The cross-sectional morphology of the elongated fibres generated *in vitro* was indistinguishable from that of the short yeast seeds, as visualized by electron microscopy. Electron diffraction of the long fibres showed the 4.7 Å spacing characteristic of the cross-beta structure of amyloids. The fact that the amyloid fibres nucleated *in vitro* propagate the strain-specific infectivity of the yeast seeds implies that the heritable infor-

mation of distinct prion strains must be encoded by different, self-propagating cross-beta folding patterns of the same prion protein.

The cellular propagation of $[PSI]$ requires an N-terminal prion-inducing fragment of the Sup35 protein⁶. Overexpression of a fusion construct, Sup35(1–61)–GFP, consisting of green fluorescent protein (GFP) fused to the carboxy terminus of the first 61 amino-acid residues of Sup35, labels numerous small, fast-moving fluorescent particles in essentially every cell harbouring the $[PSI]$ element ($[PSI^+]$ cell), but produces diffused cytoplasmic labelling in $[psi^-]$ cells⁷.

We asked whether the labelled particles observed in $[PSI^+]$ cells could transmit strain-specific phenotypes. Our experimental scheme for testing the propagation of prion infectivity is diagrammed in Fig. 1a. To facilitate the purification of $[PSI]$ particles,

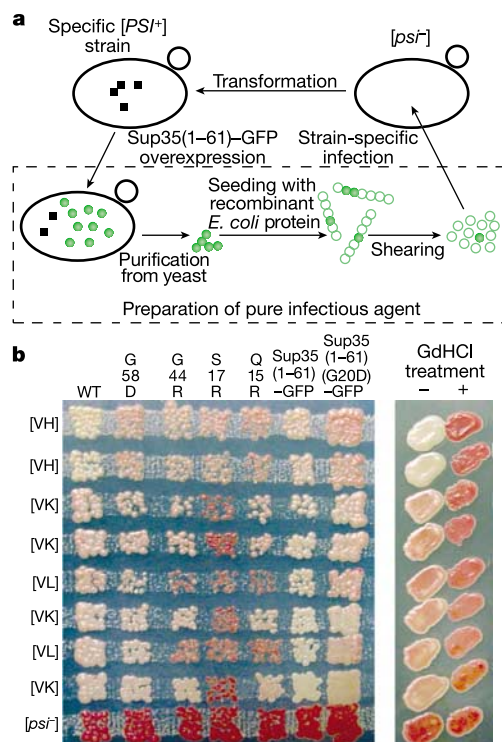


Figure 1 Proof of strain-specific infectivity of $[PSI]$ prion aggregates. **a**, Experimental scheme to demonstrate that Sup35 amyloid particles transmit the $[PSI]$ strain phenotype. Filled black squares, endogenous full-length Sup35 aggregates; filled green circles, Sup35(1–61)–GFP-labelled particles; open green circles, Sup35(1–61)–GFP aggregates derived from *E. coli*. **b**, Strain typing. A set of yeast centromere-based plasmids encoding the wild type and single mutations of the Sup35 protein (labelled WT, G58D (that is, Gly to Asp mutation at amino acid 58), G44R, S17R and Q15R) as well as yeast 2 μ -based multi-copy plasmids encoding the wild-type Sup35(1–61)–GFP or the G20D mutation⁷ were introduced into the fused cell by mating with testers that carry the plasmids. The strain type (labelled on the left) is determined by the distinctive pattern of colour responses to the full-length Sup35 proteins. For $[VH]$ the colour responses are (from lane 1 (WT) to lane 5 (Q15R)) – (changing little with respect to WT (self reference)), + (a substantial change), $\pm$ (an intermediate change), +, +, +; for $[VK]$ –, –, –, +, +; for $[VL]$ –, –, –, +, +; and for $[psi^-]$ –, –, –, –, –. Selected cells with Sup35(1–61)–GFP and Sup35(1–61)(G20D)–GFP plasmids were observed further by fluorescence microscopy for strain-specific fluorescent labelling⁷. All three $[PSI]$ strains were cured by passage through media containing 4 mM guanidine hydrochloride (right panel: $[PSI^+]$, white, pink or dark pink; $[psi^-]$, red; see Supplementary Information). In this figure, we strain-typed colonies transformed with infectious yeast particles. $[VH]$ particles were used for the top two rows. $[VK]$ particles were used for the third and fourth rows. The fifth to eighth rows were from a $[VL]$ preparation containing particles of both $[VK]$ and $[VL]$ types. The bottom row is a buffer control.

a C-terminal 10-amino-acid affinity tag, Strep-tag (II), was attached to the fusion protein. Sup35(1–61)–GFP–Strep(II)-labelled particles were isolated from yeast harbouring three different $[PSI]$ strains⁷, designated $[VH]$, $[VK]$ and $[VL]$, and further purified by StrepTactin affinity chromatography⁸. The prion strains could be distinguished by their characteristic responses to a panel of Sup35 mutants⁷. Four of the twelve Sup35 mutants tested in ref. 7 were used for strain typing (Fig. 1b).

An efficient cell fusion method was designed to introduce $[PSI]$ particles into yeast. Two populations of $[psi^-]$ spheroplasts with complementary genetic markers were mixed with purified particles and induced to fuse by polyethylene glycol treatment. The $[PSI]$ status and strain type of the fused cells, which were selected by the presence of both genetic markers, were then determined (Fig. 1b). As detailed in Supplementary Table S1, more than 10% of the complementary fused cells were strain-specifically transformed by $[VH]$ (30H/246; that is, 30 $[VH]$ and 216 $[psi^-]$ out of 246 randomly selected colonies) and $[VK]$ (76K/174) particles; however, in $[VL]$ transformants, all three $[PSI]$ strains were observed (3H, 11K, 52L/248). Overexpression of Sup35(1–61)–GFP–Strep(II) in yeast did not affect the faithful propagation of $[VH]$ and $[VK]$, but resulted in the appearance of $[VH]$, $[VK]$ and $[psi^-]$ colonies in overnight $[VL]$ cultures with frequencies of approximately 0.2%, 10% and 20%, respectively. A control plasmid without the fusion construct did not affect the propagation of $[VL]$. The occurrence of $[VH]$ and $[VK]$ strains in $[VL]$ transformants can therefore be traced to the instability of original $[VL]$ cultures when the truncated Sup35 protein was overexpressed. Strain competition experiments, similar to those carried out with the mammalian prion⁹, demonstrated that $[VH]$ particles can dominate $[VK]$ and $[VL]$ particles; whereas $[VK]$ particles can dominate $[VL]$ particles (Supplementary Table S2).

We did not obtain $[PSI^+]$ transformants when either similarly obtained preparations from $[psi^-]$ cells or soluble Sup35(1–61)–GFP (purified from $[psi^-]$ cells) was used. We also isolated Sup35(1–61)–GFP-0 particles lacking the Strep(II) affinity tag from yeast, and fractionated the preparations by StrepTactin chromatography. Although $[PSI]$ infectivity remained in the flow-through fractions, the infectivity in the column eluates was diminished. These results (Supplementary Table S1) established that the observed $[PSI]$ infectivity was associated with the Sup35(1–61)–GFP labelled particles, rather than with other fortuitously co-purified factors.

Table 1 Propagation of $[PSI]$ activity *in vitro*

Prion strain	Sup35(1–61)–GFP 1/160 dilution	Ure2(1–81)–GFP 1/160 dilution	Buffer E 1/160 dilution	Buffer E 1/80 dilution
Propagation of $[PSI]$ using unfractionated extract from <i>E. coli</i>				
$[VH]$ seeds	39H (248)	0 (248)	0 (248)	0 (248)
$[VK]$ seeds	12K (248)	0 (235)	0 (192)	1K (248)
$[VL]$ seeds	4L (248)	0 (248)	0 (248)	0 (248)
Buffer (mock)	0 (248)	0 (231)	0 (124)	0 (248)
Propagation of $[PSI]$ activity using purified His ₅ –Sup35(1–61)–GFP–Strep(II)				
Prion strain	2nd propagation 1/100 dilution	4th propagation 1/400 dilution	5th propagation 1/800 dilution	Buffer E 1/100 dilution
$[VH]$ seeds	1H (124)	11H, 1K (124)	23H (124)	0 (124)
$[VK]$ seeds	12K (124)	11K (124)	17K (124)	0 (124)
$[VL]$ seeds	0 (124)	2L (124)	3L, 3K (124)	0 (124)
Buffer (mock)	0 (124)	2K (124)	1K (124)	0 (124)

Numbers and strain types of transformed yeast (H, $[VH]$; K, $[VK]$; L, $[VL]$) are listed. Numbers of randomly screened colonies are in parentheses. $[PSI]$ seeds purified from yeast in buffer E or buffer E alone (first column) were added to extract of *E. coli* expressing His₅–Sup35(1–61)–GFP–Strep(II) or His₅–Ure2(1–81)–GFP–Strep(II) (top) or to purified His₅–Sup35(1–61)–GFP–Strep(II) (bottom). The dilution factor with respect to the original seed in each sample is indicated. Seeds directly diluted in buffer E were used as controls. The same batch of spheroplasts was used for experiments described in the same row. Transformation by mock-seeded samples (bottom row) probably resulted from *de novo* generation of $[VK]$ in the purified protein solution. Spontaneous aggregation of the recombinant protein, after extended incubation, to form infectious amyloid aggregates was reproducible.

Enzymatic modification experiments further confirmed the proteinaceous nature of the infectious particles. Whereas extended incubation of the [PSI] particles with protease K resulted in complete loss of infectivity, strain-specific propagation of infectivity was unaffected by RNase A treatment (Supplementary Table S3).

De novo generation of strain-specific infectivity was demonstrated in a set of key experiments where the infectious yeast [PSI] particles were used to nucleate the assembly of a bacterially derived Sup35(1–61)–GFP construct (Table 1). Yeast [PSI] particles were diluted 40-fold in whole-cell extracts of *Escherichia coli*, which were engineered to overexpress His₅–Sup35(1–61)–GFP–Strep(II) (that is, Sup35(1–61)–GFP–Strep(II) with an N-terminal affinity tag containing five histidine residues to facilitate purification of the protein on a Ni-NTA column). The reactions were incubated at room temperature in an undisturbed state for 24 h and then briefly sonicated. One-half of the content in the reactions was added to an equal volume of the *E. coli* extract and incubated for another 24 h. We performed the identical sonication, dilution and incubation once more and collected the aggregates by ultracentrifugation. The resulting aggregates contained about 160-fold dilution of the original infectious material. The aggregates were resuspended and sonicated in the same buffer as the original seeding particles and were used to transform yeast. Although the newly nucleated aggregates retained the strain-specific infectivity, no activity of the original seeding material, diluted 160-fold and then sonicated, could be detected (Table 1).

Propagation of [PSI] activity requires the recombinant Sup35(1–61) fragment. This was demonstrated by parallel control experiments where the Sup35(1–61) fragment in the *E. coli* cell extract was replaced with a prion-inducing fragment (amino-acid residues 1–81) of the Ure2 protein, which is essential for the propagation of the yeast prion [URE3]^{5,10}. No [PSI] activity was detected from these reactions (Table 1). Because the molecular content of the Ure2(1–81)-containing cell extract mimicked the Sup35(1–61) extract except for the prion-inducing fragments, it is evident that the Sup35(1–61) fragment was necessary for the *in vitro* propagation of [PSI] infectivity.

We next asked whether Sup35(1–61)–GFP alone was sufficient for the propagation of [PSI] infectivity. Soluble His₅–Sup35(1–61)–GFP–Strep(II) was purified from the *E. coli* extract by Ni-NTA affinity chromatography, and was used for the same serial seeding experiments as described above. Although the infectivity of the original seeds was already below detection at 100-fold dilution, the strain-specific infectivity of seeded material nearly always increased after each dilution–propagation cycle (Table 1). The newly assembled aggregates of different [PSI] strains were analysed by SDS-polyacrylamide gel electrophoresis (PAGE) (Supplementary Fig. S1) and isoelectric focusing (pH range 3–10, data not shown). The dissociated aggregates exhibited identical gel mobilities to that of the original soluble His₅–Sup35(1–61)–GFP–Strep(II) fusion protein. We did not detect any nucleic acid in purified His₅–

Sup35(1–61)–GFP–Strep(II) by gel electrophoresis (ethidium bromide staining; see Supplementary Methods), neither did the presence of RNase A in the seeding reactions influence the propagation of strain-specific infectivity (Supplementary Table S3b). Taken together, these data indicated that the *in vitro* replication of [PSI] strains only requires the Sup35(1–61)–GFP construct.

We next demonstrated the amyloid-like nature of [PSI] strains (Fig. 2). Numerous rod-shaped particles with variable lengths (20–150 nm) and a uniform width of about 170–180 Å were seen by negative staining electron microscopy in the infectious yeast preparations (Fig. 2a, top row). These structures were the only entities recognized by a GFP-specific antibody (Fig. 2a, middle row). Because the infectivity was associated with Sup35(1–61)–GFP in yeast preparations (see above), these rods must be the [PSI] determinant. When incubated with purified His₅–Sup35(1–61)–GFP protein, the short rods grew to lengths greater than 1 µm (Fig. 2a, bottom row, and panels e, f). As visualized by negative staining and cryoelectron microscopy, the fibre morphology of the long rods was indistinguishable from that of the short infectious yeast particles used for nucleation. The measured diameters of

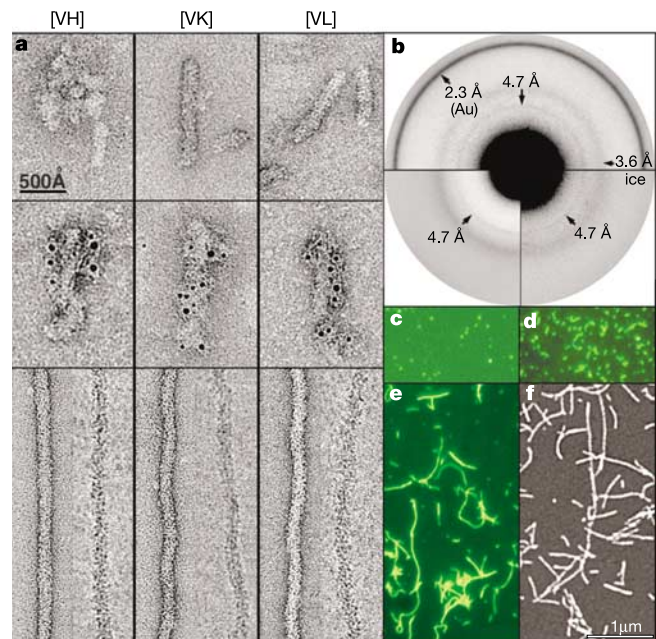


Figure 2 [PSI] amyloid fibres. **a**, Electron micrographs of the three [PSI] strain protein aggregates. The top row shows negatively stained Sup35(1–61)–GFP-labelled yeast particles. The middle row shows immunoelectron micrographs of the yeast particles, where particles decorated by a mouse anti-GFP monoclonal antibody were probed with a gold-conjugated anti-mouse IgG antibody. The bottom row shows electron micrographs of negatively stained (left) and ice-embedded (right) Sup35(1–61)–GFP amyloid fibres whose growth was nucleated by the yeast particles. At this resolution, definite differences cannot be established, although we observed fewer curved fibres in [VH] samples than in [VK] and [VL] samples. The morphology of these negatively stained fibres is similar to that of recombinant Ure2(1–65)–GFP fibres²⁶. **b**, Electron diffraction patterns of ice-embedded, unoriented fibre specimens: [VH] (top half), [VK] (bottom left) and [VL] (bottom right), showing the characteristic 4.7 Å spacing of the cross-beta structure of amyloids. The 2.3 Å ring is from evaporated gold, used for calibration. The diffuse ring at approximately 3.6 Å spacing is from vitreous ice. The size of the central inelastic scattering maximum depends on the contrast used to display the sharp 4.7 Å ring. A similar X-ray diffraction pattern was reported from spontaneously formed Sup35 fibres from *E. coli*²⁷. **c–e**, Evolution of fibre morphologies in a typical seeding cycle described in Table 1. The fluorescence micrographs show yeast-derived [VK] particles (**c**) and growth after 48 h incubation with *E. coli*-derived Sup35(1–61)–GFP (**e**). The long fibres are sheared by sonication to release numerous smaller seeds (**d**). **f**, Electron micrograph of elongated fibres at a magnification similar to **e**.

Table 2 Increase of [PSI] infectivity by sonic disruption

Prion strain	+Incubation +sonication	+Incubation –sonication	–Incubation +sonication	–Incubation –sonication
[VH] seeds	41H (248)	17H (248)	0 (248)	3H (248)
[VK] seeds	77K (248)	14K (248)	17K (248)	20K (248)
[VL] seeds	16L, 1K (248)	2L (248)	2L, 1K (248)	1L, 1K (248)
Buffer (mock)	0 (248)	0 (248)	ND	0 (248)

Numbers and strain types of transformed yeast are listed in the same fashion as in Table 1. Pre-sonicated [PSI] seeds in buffer E as well as buffer E control (first column) were added to highly purified recombinant His₅–Sup35(1–61)–GFP–Strep(II) protein, either immediately before yeast transformation (–incubation) or allowing incubation at room temperature for 48 h before transformation (+incubation). Before transformation, the mixtures were divided into two equal portions. One was treated with brief sonication (+sonication), and the other was left untreated (–sonication). The same batch of spheroplasts was used for all experiments. The moderate increase of infectivity after 48 h incubation in untreated [VH] sample (+incubation, –sonication) but not in untreated [VK] or [VL] samples is reproducible with different batches of seeds. This fact seems to indicate that [VH] fibrils are more fragile. ND, not determined.

negatively stained long fibres are $170 \pm 11 \text{ \AA}$ for [VH], $177 \pm 9 \text{ \AA}$ for [VK] and $177 \pm 11 \text{ \AA}$ for [VL]. Electron diffraction of the ice-embedded long fibres showed the $4.7 \text{ \AA}$ spacing characteristic of the cross-beta structure (Fig. 2b).

The long fibres could be broken into numerous short fibrils by sonication as monitored by fluorescence microscopy (Fig. 2d, e) and electron microscopy (not shown). The multiplication of fibrils thus provided an explanation for the observed increase of infectivity in the serial seeding experiments described in Table 1. We tested this surmise by comparing the infectivity of seeded aggregates with and without sonication. Yeast-derived pre-sonicated aggregates were used to seed the assembly of *E. coli*-derived highly purified His₅-Sup35(1–61)-GFP-Strep(II) at room temperature without agitation. After 48 h, each reaction was divided into two equal portions. One was subjected to brief sonication, while the other was left undisturbed. These samples were then used to transform yeast. As controls, the same amount of the yeast aggregates was added in exactly the same way to the same batch of unpolymerized *E. coli* protein without incubation, and then treated with or without sonication before assaying for infectivity. For every [PSI] strain, brief sonication increases the efficiency of yeast transformation for the incubated, seeded samples compared with the controls (Table 2).

The enhanced infectivity of the fragmented amyloid fibres can be attributed to the increase in the number of particles after sonication, although fragmentation may also facilitate transformation if the shorter filaments are more efficiently assimilated on fusion of the yeast cells. The sheared fibres were on average longer than the original yeast particles as judged by fluorescence (Fig. 2c, d) and electron microscopy, yet they were more numerous and thus collectively more infectious. This point was illustrated further by the serial seeding experiments described in Table 1 where sheared fibres in each growth cycle were of similar sizes, but the infectivity of the samples almost always increased after each cycle. In yeast the propagation of [PSI] depends on Hsp104 activity¹¹. The Hsp104 protein is believed to fragment the Sup35 aggregates in [PSI⁺] cells, thus allowing the replication and propagation of [PSI]¹². Our experiment, correlating mechanical shearing of fibres with increase in infectivity *in vitro*, seems to mirror the proposed *in vivo* mechanism.

The data presented so far, namely the nucleated growth of strain-specific [PSI] infectivity in solutions of purified recombinant protein, strongly support the protein-only nature of [PSI] strains as well as the adequacy of the Sup35(1–61) construct in propagating strain-specific biological activity *in vitro* (Fig. 1a). Although catalytic amounts of yeast particles were used as seeds to encode strain-type information, their presence was successively reduced by serial seeding in the experiments presented in Table 1. To corroborate further the 'protein-only' assertion, we carried out an independent set of transformation experiments similar to those of ref. 13 using bacterially derived Sup35(1–61)-GFP fibres, which were spontaneously aggregated from pure protein solutions.

Two different aggregates were prepared without seeding by extended incubation of *E. coli*-derived His₅-Sup35(1–61)-GFP-Strep(II). Sample I, obtained from buffer E (100 mM TrisHCl, 1 mM EDTA, 2.5 mM desthiobiotin, pH 8.0) at 4 °C, was concentrated by centrifugation and resuspension in the same buffer. Sample II was obtained from buffer B (20 mM TrisHCl, 100 mM NaCl, 500 mM imidazole, pH 7.6) at 22 °C, collected by centrifugation, and resuspended in buffer E. Electron microscopy revealed amyloid fibres with similar morphology in both samples (Supplementary Fig. S2). The samples were checked by SDS-PAGE (Supplementary Fig. S1) and isoelectric focusing. The dissociated samples appeared identical with regard to their monomer size and the electric charge compared to those of the soluble His₅-Sup35(1–61)-GFP-Strep(II). Yeast colonies transformed by sample I exhibited the [VH] strain phenotype (38H/248, 20H/248 for 5/8 dilution) and colonies transformed by sample II displayed the [VK] pheno-

type (2K/124, 2K/248, same sample; the same spheroplast preparations were used for the latter and the 5/8 dilution of sample I). The observed [PSI] induction cannot merely be attributed to the fact that overexpression of the Sup35 protein induces [PSI] *de novo* in yeast^{14,15}—such a procedure lacks prion-strain specificity¹⁶. Instead, the strain-specific transformation indicates the transmission of specific genetic information from the two samples, thus qualifying spontaneously aggregated Sup35 amyloid fibres as bona fide infectious agents. As a further control, we reconfirmed the lack of *de novo* [PSI] generation in our experimental yeast genetic background when His₅-Sup35(1–61)-GFP was overexpressed from a multi-copy yeast plasmid (see Supplementary Information for details). These observations demonstrate that [PSI] infectivity can also arise from self-assembly of the pure *E. coli* protein construct, which was never in contact with yeast cell extract.

Our results demonstrate that [PSI] strains are propagated by infectious amyloid aggregates of the Sup35 prion-forming domain, thus resolving current ambiguity regarding the molecular nature of [PSI]^{13,16,17} and the mechanism for its strain variations¹⁸. 'Protein-only' [PSI] implies that the prion strain differences—distinguished by testing their responses with a panel of mutant Sup35 proteins—must be structural. Although the *in vivo* propagation of [PSI] after infection probably involves complex cellular interactions¹⁹, our data clearly show that the first 61 amino-acid residues of the Sup35 protein are sufficient for encoding the strain-specific infectivity. The challenge now is to characterize the structural differences in the cross-beta folding patterns of the same 61-amino-acid fragment that encodes the heritable information for different prion strains. □

Methods

Plasmids and yeast genetic backgrounds

The construction of YE_p-CUP1-Sup35(1–61)-GFP-Strep(II)-T and YE_p-CUP1-His₅-Sup35(1–61)-GFP-Strep(II)-T is described in Supplementary Information. The plasmids YC_p-I-SUPFs (URA3) have been described previously⁷.

The *E. coli* plasmids pHis-Sup35(1–61)-GFP-Strep(II) and pHis-Ure2(1–81)-GFP-Strep(II) were constructed on the T7 expression vector pMW172 (ref. 20; see also Supplementary Information). Yeast strains were derived from 5V-H19 (MAT_a SUQ5 *ade2-1*(UAA) *can1-100 leu2-3,112 ura3-52*)⁶. For [PSI] transformation, g- α -5V-H19 (MAT_a SUQ5 *ade2-1*(UAA) *can1-100 leu2-3,112 ura3-52* [*psi*[−]] [*pin*[−]]) and g- α -5V-H19 Δ HIS4 [YCplac111] (g- α -5V-H19 *his4::loxP-kanMX-loxP* [YCplac111 (LEU2)]) were used. We used their [PSI⁺] [PIN⁺] derivatives for strain competition. *Escherichia coli* strain BLR (DE3)/pLysS (ref. 21) was purchased from Novagen.

Purification of [PSI] particles

[PSI] particles were purified from cell lysates of 5V-H19 (MAT_a/MAT_a [PIN⁺] derivatives) transformed with YE_p-CUP1-Sup35(1–61)-GFP-Strep(II)-T (Supplementary Information). Lysates were placed on top of 30% (g ml^{−1}) sucrose (in buffer A (20 mM TrisHCl, 100 mM NaCl, 5 mM imidazole, pH 7.6)) for 2 h sedimentation at 200,000g. Aggregates were re-suspended in 2 ml buffer A plus 50 μ g avidin, sonicated briefly, and then loaded to 1 ml StrepTactin columns (IBA). The columns were washed at room temperature with 10 vol. of buffer W (100 mM TrisHCl, 1 mM EDTA, pH 8.0), 15 vol. of buffer A, and then eluted by 6 $\times$ 0.5 ml buffer E (buffer W plus 2.5 mM desthiobiotin). The third and fourth elution fractions contained most of the [PSI] particles and were collected. The first 1 ml wash was used as unbound fractions for transformation experiments (Supplementary Table S1). Soluble Sup35(1–61)-GFP-Strep(II) was similarly purified from a [*psi*[−]] [*pin*[−]] derivative, except that the supernatant above the sucrose cushion was used for subsequent StrepTactin chromatography.

Escherichia coli protein preparation and *in vitro* seeding

Bacterial protein overexpression was induced by 1 mM isopropyl- β -D-thiogalactopyranoside at an optical density of 0.6 at 595 nm (LB media, 37 °C) in BLR (DE3)/pLysS cells transformed with pHis-Sup35(1–61)-GFP-Strep(II) or pHis-Ure2(1–81)-GFP-Strep(II). *Escherichia coli* cell extracts were prepared 3 h after induction by removing debris (12,000g sedimentation, twice) from cell lysates, which were obtained by sonication of 0.5 g ml^{−1} cell suspension in buffer A plus 1 mM phenylmethylsulphonyl fluoride.

Cell extracts were loaded onto 7 ml Ni-NTA affinity columns (Qiagen) at 4 °C. The columns were washed extensively with buffer A. Protein was eluted by buffer B (buffer A plus 500 mM imidazole) and stored immediately at −80 °C. Samples thus prepared were used for seeding experiments as purified protein. For experiments in Table 2, a sample was further purified by StrepTactin affinity chromatography and used immediately. Typical recombinant protein concentrations for seeding experiments were between 10 and 40 μ M. In this report 'ultracentrifugation' refers to 80,000g sedimentation on top of 1 ml 30%

(g ml⁻¹) sucrose cushion for 2 h and 'brief sonication' indicates a 10 s sonic disruption with 6 W power output. Unpolymerized protein solutions were checked by fluorescence microscopy for lack of visible aggregation before seeding experiments.

Yeast transformation by [PSI⁺] particles

Yeast spheroplasts were prepared from g- α h-5V-H19 and g- α h-5V-H19 Δ HIS4 [YCplac111] in buffer Z1 (1.2 M sorbitol, 10 mM TrisHCl (pH 7.5), 30 mM CaCl₂) (Supplementary Information). Fifty microlitres of each spheroplast suspension were mixed with 8 μ l solution containing prion particles, incubated at 22 °C for 15 min, and then combined. One millilitre of buffer Z2 (20% (w/v) polyethylene glycol 3350, 10 mM TrisHCl (pH 7.5), 30 mM CaCl₂) was added to the mixture. After 15 min incubation at 22 °C, spheroplasts were collected and allowed to recover in 150 μ l Z3 medium (1 M sorbitol, 30 mM CaCl₂, one-third strength YPAD²², 7 mg l⁻¹ uracil, 7 mg l⁻¹ histidine, 33 mg l⁻¹ leucine) at 30 °C for 30 min. Top agar (SC-LEU, HIS, with 1.2 M sorbitol, 8 ml) was then added to the spheroplasts to plate on top of SC-LEU, HIS agar²² containing 1.2 M sorbitol. Karyogamy in the 5V-H19 background was efficient^{23,24}. A control transformation by pure sample buffer was included for every set of experiments. We never obtained a transformed colony from these controls.

Strain typing

Randomly selected colonies were transferred from top agar to SC-LEU, HIS plates and incubated at 30 °C for 5 days to allow colony colour to fully develop. All white, pink and dark pink colonies were tested for [PSI⁺] by mating on SC-URA, LEU, HIS plates with a set of *MATa/MATa* [*psi*⁻] testers, transformed with YEp-CUP1-N1-GFP-T, YEp-CUP1-N1(G20D)-T, YCp-I-Sup35F, YCp-I-Sup35F(G58D), YCp-I-Sup35F(G44R), YCp-I-Sup35F(S17R), and YCp-I-Sup35F(Q15R). The strain types were determined by characteristic colour responses⁷ (Fig. 1b). See Supplementary Information for more details.

Electron microscopy

Samples for immunoelectron microscopy were prepared by the procedure of Wischik *et al.*²⁵ with a 1:2,000 dilution of the anti-GFP antibody JL8 (Clontech), followed by a 1:5 dilution of goat anti-mouse IgG, cross-linked to 6 nm gold particles (EMS), and then stained with 1% (w/v) uranyl acetate.

Ice-embedded samples were used for cryoelectron microscopy and electron diffraction. Images were collected at $\times 45,000$ magnification under low dose conditions.

Received 11 November 2003; accepted 2 February 2004; doi:10.1038/nature02391.

1. Prusiner, S. B. Prions. *Proc. Natl Acad. Sci. USA* **95**, 13363–13383 (1998).
2. Dickinson, A. G. & Outram, G. W. Genetic aspects of unconventional virus infections: the basis of the virino hypothesis. *Ciba Found. Symp.* **135**, 63–83 (1988).
3. Weissmann, C. A 'unified theory' of prion propagation. *Nature* **352**, 679–683 (1991).
4. Cox, B. S., Tuite, M. F. & McLaughlin, C. S. The *psi* factor of yeast: a problem in inheritance. *Yeast* **4**, 159–178 (1988).
5. Wickner, R. B. [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* **264**, 566–569 (1994).
6. Ter-Avanesyan, M. D., Dagkesamanskaya, A. R., Kushnirov, V. V. & Smirnov, V. N. The SUP35 omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [PSI⁺] in the yeast *Saccharomyces cerevisiae*. *Genetics* **137**, 671–676 (1994).
7. King, C.-Y. Supporting the structural basis of prion strains: induction and identification of [PSI⁺] variants. *J. Mol. Biol.* **307**, 1247–1260 (2001).
8. Skerra, A. & Schmidt, T. G. Use of the Strep-Tag and streptavidin for detection and purification of recombinant proteins. *Methods Enzymol.* **326**, 271–304 (2000).
9. Dickinson, A. G. *et al.* Extraneal competition between different scrapie agents leading to loss of infectivity. *Nature* **253**, 556 (1975).
10. Maddelein, M. L. & Wickner, R. B. Two prion-inducing regions of Ure2p are nonoverlapping. *Mol. Cell. Biol.* **19**, 4516–4524 (1999).
11. Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Liebman, S. W. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [PSI⁺]. *Science* **268**, 880–884 (1995).
12. Kushnirov, V. V. & Ter-Avanesyan, M. D. Structure and replication of yeast prions. *Cell* **94**, 13–16 (1998).
13. Sparrer, H. E., Santoso, A., Szoka, F. C. & Weissman, J. S. Evidence for the prion hypothesis: induction of the yeast [PSI⁺] factor by *in vitro*-converted Sup35 protein. *Science* **289**, 595–599 (2000).
14. Chernoff, Y. O., Derkach, I. L. & Inge-Vechtomov, S. G. Multicopy SUP35 gene induces *de-novo* appearance of *psi*-like factors in the yeast *Saccharomyces cerevisiae*. *Curr. Genet.* **24**, 268–270 (1993).
15. Derkach, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G. & Liebman, S. W. Genesis and variability of [PSI⁺] prion factors in *Saccharomyces cerevisiae*. *Genetics* **144**, 1375–1386 (1996).
16. Liebman, S. W. Progress toward an ultimate proof of the prion hypothesis. *Proc. Natl Acad. Sci. USA* **99**, 9098–9100 (2002).
17. Tuite, M. F. & Cox, B. S. Propagation of yeast prions. *Nature Rev. Mol. Cell. Biol.* **4**, 878–890 (2003).
18. Uptain, S. M., Sawicki, G. J., Caughey, B. & Lindquist, S. L. Strains of [PSI⁺] are distinguished by their efficiencies of prion-mediated conformational conversion. *EMBO J.* **20**, 6236–6245 (2001).
19. Liu, J.-J., Sondheimer, N. & Lindquist, S. L. Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion [PSI⁺]. *Proc. Natl Acad. Sci. USA* **99**, 16446–16453 (2002).
20. Way, M., Pope, B., Gooch, J., Hawkins, M. & Weeds, A. G. Identification of a region in segment 1 of gelsolin critical for actin binding. *EMBO J.* **9**, 4103–4109 (1990).
21. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89 (1990).
22. Sherman, F. Getting started with yeast. *Methods Enzymol.* **194**, 3–21 (1991).
23. Harashima, S., Takagi, A. & Oshima, Y. Transformation of protoplasted yeast cells is directly associated with cell fusion. *Mol. Cell. Biol.* **4**, 771–778 (1984).
24. Rose, M. D., Price, B. R. & Fink, G. R. *Saccharomyces cerevisiae* nuclear fusion requires prior activation by alpha factor. *Mol. Cell. Biol.* **6**, 3490–3497 (1986).

25. Wischik, C. *et al.* Structural characterization of the core of the paired helical filament of Alzheimer disease. *Proc. Natl Acad. Sci. USA* **85**, 4884–4888 (1988).
26. Baxa, U. *et al.* Architecture of Ure2p prion filaments. *J. Biol. Chem.* **278**, 43717–43727 (2003).
27. Serio, T. R. *et al.* Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* **289**, 1317–1321 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. L. D. Caspar for support, suggestions and assistance with the manuscript, and C. Long and B. Patel for reviewing the manuscript. This work was supported by an NIH grant to D. L. D. Caspar and an NIH Research Service Award to C.-Y.K.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.-Y.K. (chihyen@sb.fsu.edu).

Conformational variations in an infectious protein determine prion strain differences

Motomasa Tanaka¹, Peter Chien^{1,2}, Nariman Naber³, Roger Cooke³ & Jonathan S. Weissman^{1,2}

¹Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, ²Graduate Group in Biophysics, ³Department of Biochemistry and Biophysics, University of California-San Francisco, San Francisco, California 94143, USA

A remarkable feature of prion biology is the strain phenomenon wherein prion particles apparently composed of the same protein lead to phenotypically distinct transmissible states^{1–4}. To reconcile the existence of strains with the 'protein-only' hypothesis of prion transmission, it has been proposed that a single protein can misfold into multiple distinct infectious forms, one for each different strain^{1–3,5}. Several studies have found correlations between strain phenotypes and conformations of prion particles^{6–10}; however, whether such differences cause or are simply a secondary manifestation of prion strains remains unclear, largely due to the difficulty of creating infectious material from pure protein^{3,5}. Here we report a high-efficiency protocol for infecting yeast with the [PSI⁺] prion using amyloids composed of a recombinant Sup35 fragment (Sup-NM). Using thermal stability and electron paramagnetic resonance spectroscopy, we demonstrate that Sup-NM amyloids formed at different temperatures adopt distinct, stably propagating conformations. Infection of yeast with these different amyloid conformations leads to different [PSI⁺] strains. These results establish that Sup-NM adopts an infectious conformation before entering the cell—fulfilling a key prediction of the prion hypothesis⁵—and directly demonstrate that differences in the conformation of the infectious protein determine prion strain variation.

The yeast prion state [PSI⁺]¹¹ results from self-propagating aggregation of the translation termination factor protein Sup35, leading to a nonsense suppression phenotype. In yeast containing a nonsense mutation in the *ade1* gene, [PSI⁺] colonies are white or pink and grow on media lacking adenine, whereas [*psi*⁻] colonies are red and require adenine. Transient overexpression of the Sup35 amino-terminal, glutamine/asparagine-rich (prion) domain leads to Sup35 aggregation and *de novo* appearance of [PSI⁺]. *In vitro*, this prion domain forms self-seeding amyloid fibres^{12,13}. As with mammalian prions and other yeast prions, [PSI⁺] exhibits a range of heritable phenotypic strain variants¹⁴ that differ in mitotic

stability¹⁴, dependence on the cellular chaperone machinery¹⁵ and in the solubility and activity of Sup35 (refs 14, 16–18), leading to differences in the *adel* colour phenotype. Sup35 aggregates obtained from different $[PSI^+]$ strains also vary in their ability to seed polymerization of purified protein *in vitro*¹⁸. In addition, $[PSI^+]$ strains can have a major role in determining the specificity of prion transmission^{9,10,19}. This link between prion strains and transmission barriers is likely to be general, as the ability of mammalian prions to cross species barriers also differs substantially among mammalian prion strains^{2,4}.

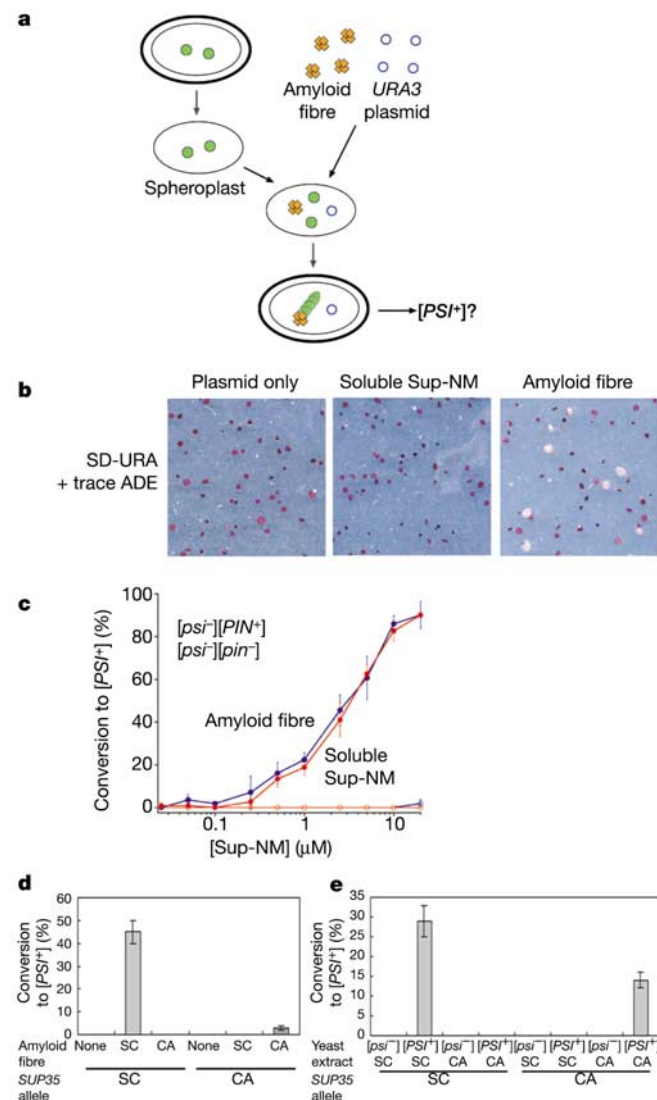


Figure 1 Induction of the $[PSI^+]$ prion by *in-vitro*-converted Sup-NM amyloid fibres. **a**, Schematic diagram of the transformation procedure. The $[PSI^+]$ status is assessed by plating on SD-URA containing trace amounts of adenine. **b**, Examples of transformations using the indicated material for infections. Large white colonies represent authentic $[PSI^+]$ cells (see Supplementary Figs S2 and S3). **c**, Concentration and $[PIN^+]$ state dependence of prion infectivity. The indicated concentration of fibrillar (filled circles) or soluble (open circles) Sup-NM was transformed into isogenic $[psi^-][PIN^+]$ (blue) or $[psi^-][pin^-]$ (red) strains. Values are expressed as mean $\pm$ standard deviation. **d**, Species-specific infection by *in-vitro*-converted amyloid fibres. SC and CA refer to prion domains derived from *S. cerevisiae* and *C. albicans* SUP35, respectively. SUP35 allele refers to the sequence of the genomic copy of the SUP35 prion domain of the transformed yeast. **e**, Species-specific infection by crude yeast extracts. $[psi^-]CA$ and $[PSI^+]CA$ refer to *S. cerevisiae* in which the genomic SUP35 prion domain was replaced by the *C. albicans* sequence and Sup35 is in a soluble or prion form, respectively.

Here we present a high-efficiency protocol for infecting yeast with preformed prion particles. Earlier studies demonstrated that introduction by liposome fusion of bacterially produced Sup-NM (Sup35 N-terminal residues 1–254) induces $[PSI^+]$ ²⁰; however, this liposome-based infection strategy was poorly suited to test the role of prion conformations in strain diversity as only a small fraction (1–2%) of yeast were converted to $[PSI^+]$. More fundamentally, Sup-NM amyloid formation occurs after encapsulation within liposomes, making it difficult to control for or evaluate the conformation of the infectious form. To overcome these difficulties, we have developed a modified transformation protocol in which preformed Sup-NM amyloid fibres along with a *URA3*-marked plasmid are introduced into yeast spheroplasts by treatment with polyethylene glycol (PEG) (Fig. 1a). The inclusion of the *URA3* plasmid allows us to select for the small fraction of yeast that have successfully taken up material from solution. After spheroplast transformation, cells are grown on plates that are devoid of uracil and contain trace amounts of adenine. On such plates, $[psi^-]$ yeast form small, intensely red colonies, whereas $[PSI^+]$ yeast form large white colonies. Transformation of $[psi^-]$ yeast with plasmid alone or with soluble Sup-NM did not lead to detectable formation of $[PSI^+]$ colonies. In contrast, inclusion of preformed Sup-NM amyloid seeds led to significant production of large, white (*Ade*⁺) colonies (Fig. 1b). As expected for a prion, infectivity was sensitive to proteinase but not nuclease treatment (Supplementary Fig. S1). Increasing the concentration of Sup-NM fibres resulted in a dose-dependent increase in conversion to *Ade*⁺ colonies, with the

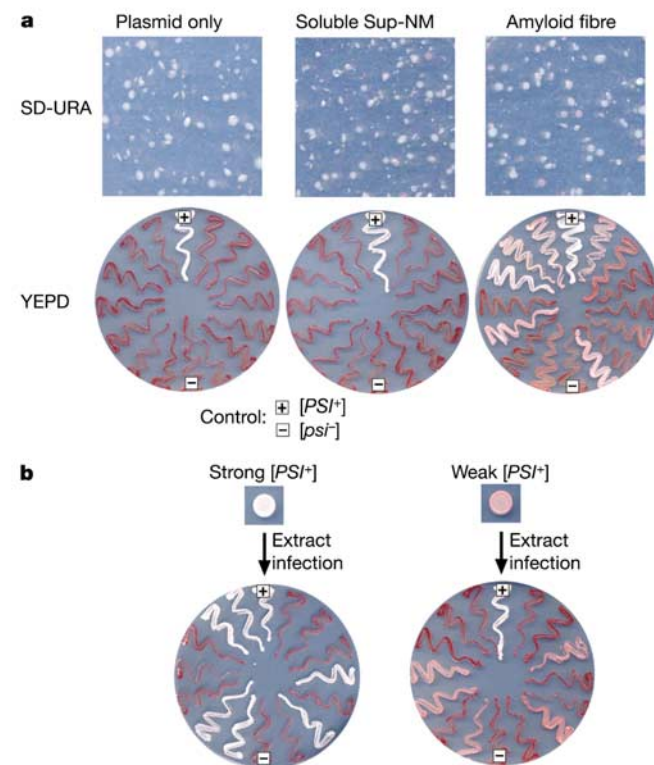


Figure 2 Generation of multiple $[PSI^+]$ strains by Sup-NM amyloid fibres converted *in vitro*. **a**, Infection of yeast without selection for the prion state. After PEG transformation with the indicated material, cells were recovered on SD-URA and randomly selected colonies were streaked on YEPD plates to identify colonies converted to $[PSI^+]$ (see Methods). Amyloid fibres converted *in vitro* induced a range of pink to white $[PSI^+]$ strains. Throughout the panels the top and bottom streaks are strong $[PSI^+]$ and $[psi^-]$ controls, respectively. **b**, Induction of prion state by yeast extracts derived from strong (white) or weak (pink) $[PSI^+]$ strains. Note that successful $[PSI^+]$ infectants show strain phenotypes of the donor strain.

fraction of Ade⁺ colonies among the Ura⁺ colonies approaching 100% at high Sup-NM concentrations (Fig. 1c). These converted Ade⁺ colonies exhibited the hallmarks of the [PSI⁺] prion state: the Ade⁺ trait was inherited in a non-mendelian manner, the trait was readily cured by transient growth on medium containing guanidine hydrochloride (GuHCl) and was associated with the formation of large Sup35 aggregates, that were pelletable by high-speed centrifugation (Supplementary Figs S2 and S3). Sup35 overexpression only induces [PSI⁺] in yeast harbouring a second prion, termed [PIN⁺]²¹ (Supplementary Fig. S4). In contrast, we found that the efficiency of infection was independent of [PIN⁺] (Fig. 1c), consistent with the requirement for [PIN⁺] in the initial conversion *in vivo* of Sup35 to the prion state but not for the subsequent propagation of [PSI⁺]. Our ability to infect [pin⁻] cells at high efficiency indicates that [PSI⁺] induction was not simply a by-product of increasing cellular prion protein levels, a possibility for which it was difficult to control for in previous protein infection experiments^{5,20,22}, rather, conversion to [PSI⁺] was a consequence of introducing Sup-NM in an infectious (prion) conformation.

We examined the specificity of our infectious particles. Whereas the prion domain of Sup35 is conserved across a range of budding yeast, [PSI⁺] 'infectivity' is often limited by transmission barriers^{23–25} analogous to the species barriers in mammalian prions. For example, prion domains of Sup35 from *Saccharomyces cerevisiae* and *Candida albicans* are both capable of forming heritable conformations but do not cross-seed each other *in vivo* or *in vitro*²³. Our infection experiments faithfully reproduced this specificity in prion transmission. *Saccharomyces cerevisiae* fibres were unable to infect yeast that had the genomic SUP35 prion domain replaced by the *C. albicans* sequence (Fig. 1d). Similarly, *C. albicans* could infect *C. albicans* yeast, albeit at low efficiency, but were unable to infect yeast that had the wild-type (*S. cerevisiae*) SUP35. Finally, crude extracts from [PSI⁺] *S. cerevisiae* or *C. albicans* could be used as a source of infectious material and displayed the same self-specificity as recombinant prions (Fig. 1e).

Induction of [PSI⁺] by Sup35 overexpression leads to a range of prion strains that can be distinguished readily by differences in their colony colour on rich medium (YEPA)¹⁴. To determine whether infection with pure Sup-NM protein led to a similar range of strains, we plated PEG-transformed cells on to medium rich in adenine (SD-URA), which allowed yeast to grow robustly regardless of the [PSI⁺] state (Fig. 2a, top). Randomly chosen colonies were then streaked on YEPA to reveal their colour phenotype, which indicated whether they remained [psi⁻] (red) or had converted to a weak [PSI⁺] (pink) or strong [PSI⁺] (white) state. As expected, mock-infected yeast or yeast infected with soluble Sup-NM yielded only red [psi⁻] colonies. In contrast, infection with preformed fibres led to a range of [PSI⁺] strains (Fig. 2a, bottom). Importantly, when we used extracts or partially purified Sup35 proteins from strong or weak [PSI⁺] strains, we exclusively observed strong or weak [PSI⁺] transformants, respectively (Fig. 2b; see also Supplementary Fig. S5). Thus, the different prion strains observed in yeast transformed with pure protein are not a consequence of the transformation procedure but rather arose from intrinsic heterogeneity in the Sup-NM prions formed *in vitro*.

Earlier studies have shown that, as with other amyloid-forming proteins^{4,26}, Sup-NM spontaneously adopts a mixture of distinct fibre types^{12,27}, suggesting that these different amyloid conformations may be responsible for the different [PSI⁺] strains generated in the above infection experiments. To explore the relationship between prion strains and amyloid conformations, we prepared Sup-NM amyloids with different conformations *in vitro*. We generated Sup-NM amyloids at different temperatures (4 °C, 23 °C and 37 °C), as previous studies have found that for some proteins polymerization temperature can influence the range of preferred amyloid conformations^{4,10}. Both resistance to thermal denaturation in the presence of SDS and limited proteolysis (Supplementary Fig. S6) indicated that fibres formed at 4 °C had markedly different physical properties than those formed at 23 °C or 37 °C (Fig. 3a). In particular, 4 °C fibres showed a relatively low melting temperature

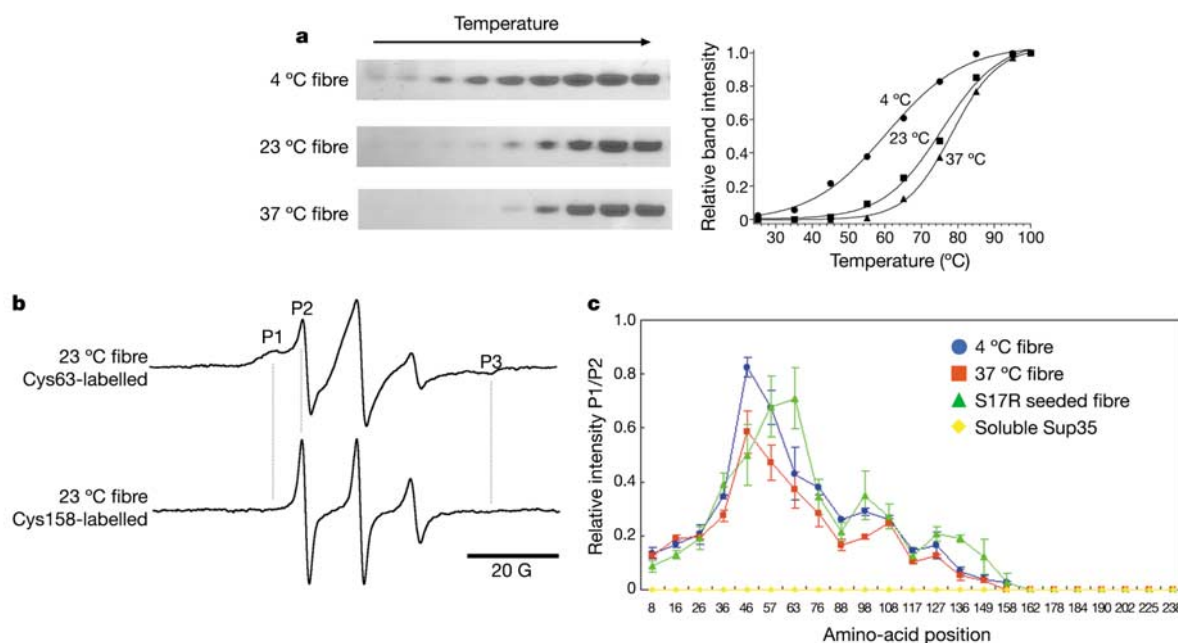


Figure 3 Different conformations of Sup-NM amyloid fibres are formed at different temperatures. **a**, Stability of 4 °C, 23 °C and 37 °C amyloid fibres as determined by SDS-PAGE. The band intensities, which reflect susceptibility of aggregates to thermal solubilization, were plotted against temperature and fitted to a sigmoidal function. **b**, EPR spectra of 23 °C amyloid fibres spin-labelled at positions 63 and 158. P1 and P3 result

from immobilized spin probes and P2 results from mobile spin probes. **c**, The relative ratio of the intensities of P1 and P2 (P1/P2) in EPR spectra was plotted against the amino-acid position for fibres formed at 4 °C or 37 °C, as well as for fibres seeded by a small amount (5%) of unlabelled S17R amyloids and soluble Sup-NM. Note that fibres are likely to represent populations of distinct amyloid subtypes²⁷.

($T_m = 62 \pm 4^\circ\text{C}$) and broad melting transition ($W = 29 \pm 6^\circ\text{C}$) compared with 23°C ($T_m = 78 \pm 3^\circ\text{C}$, $W = 18 \pm 3^\circ\text{C}$) or 37°C ($T_m = 79 \pm 2^\circ\text{C}$, $W = 12 \pm 2^\circ\text{C}$) fibres.

Electron paramagnetic resonance (EPR) spectroscopy on spin labels introduced at specific positions within Sup-NM revealed that

there are general structural features common to all Sup-NM fibres as well as localized differences characteristic of particular fibre conformations. We prepared Sup-NM mutants in which a single cysteine residue was substituted into wild-type Sup-NM (which lacks cysteines) at approximately every tenth residue. The mutants were then labelled with a cysteine-specific spin probe²⁸. For each labelled protein, we measured spectra of soluble Sup-NM and of fibres formed under various conditions. Regardless of the polymerization state, residues in the carboxy-terminal region of Sup-NM (residue 158 and beyond) displayed sharp EPR spectra indicative of highly mobile spin probes²⁹ (Fig. 3b, c). In contrast, residues in the middle of the prion domain (especially those between residues 36 and 76) in amyloid forms, but not in the unpolymerized state, showed broad peaks at low (P1) and high (P3) fields, indicative of highly immobilized spin probes²⁹. Amyloid seeds from a Sup-NM mutant (S17R) have been shown to convert wild-type Sup-NM into a conformation distinct from that produced by wild-type seeds¹⁰. We found that in S17R-seeded fibres the maximal position of the P1 peak is at residue 63 compared with the maximum seen at residue 46 in the 4°C or 37°C fibre forms (Fig. 3c). We also observed significant, site-specific differences in the intensities of the P1 peak between fibres formed at 4°C or 37°C (Fig. 3c).

Our ability to produce distinct Sup-NM amyloid forms solely by varying the polymerization temperature allowed us to test directly the role of prion conformation in determining strain properties *in vivo*. We found marked differences in the $[\text{PSI}^+]$ strains generated by prion infections depending on the conformation of the infectious material used. Fibres formed at 4°C had a high efficiency of infection with the large majority of colonies showing a strong (white) $[\text{PSI}^+]$ strain phenotype (Fig. 4a, b), whereas fibres formed at 23°C or 37°C had lower infectivity and produced almost exclusively weak (pink and/or sectorial) $[\text{PSI}^+]$ strains (Fig. 4a, b). The weak strains showed increased levels of soluble Sup35 and were more readily cured by Hsp104 overexpression (Supplementary Figs S3 and S7). In principle, these differences in prion strains could be a consequence of differences in the prion titre of 4°C and 37°C fibre preparations (for example, introduction of multiple infectious particles could promote strong strains). We excluded this possibility by demonstrating that strain phenotypes do not depend on the concentration of seed or the infection efficiency: 4°C fibres yielded strong strains even when diluted tenfold (infection rate $\sim 20\%$) and 37°C fibres yielded weak strains even when the concentration was increased tenfold (infection rate $\sim 80\%$) (Fig. 4c). Once formed, the 37°C fibre conformation is stable and can be transmitted by subsequent polymerization even at 4°C . Polymerization of Sup-NM at 4°C , induced by seeding with a small amount of 37°C fibre, leads to an amyloid with the properties of 37°C fibres; that is, a high melting temperature ($T_m = 77 \pm 4^\circ\text{C}$) and a yield of predominantly weak strains in infection experiments (Fig. 4d). Together, the above data demonstrate that a single protein can adopt multiple distinct, self-propagating (infectious) conformations and that these conformational differences underlie heritable differences in prion strains.

The existence of multiple distinct heritable strains has been difficult to explain in the context of the protein-only hypothesis of prion inheritance^{2,3}. Strains now seem to be a nearly universal feature of prions: a wide range of different strains have been found for mammalian prions, for the three known naturally occurring yeast prions, and for a variety of artificial yeast prions⁴. Cellular factors such as chaperones probably have an important role in the manifestation of strain phenotype and promoting stable strain propagation¹⁸. Nonetheless, the ubiquitous nature of strains argues that they arise from a common physical property shared by all prions rather than from a specific feature of a given prion, such as the sequence of the infectious protein. All naturally occurring prions appear to be based on self-propagating, β -sheet-rich aggre-

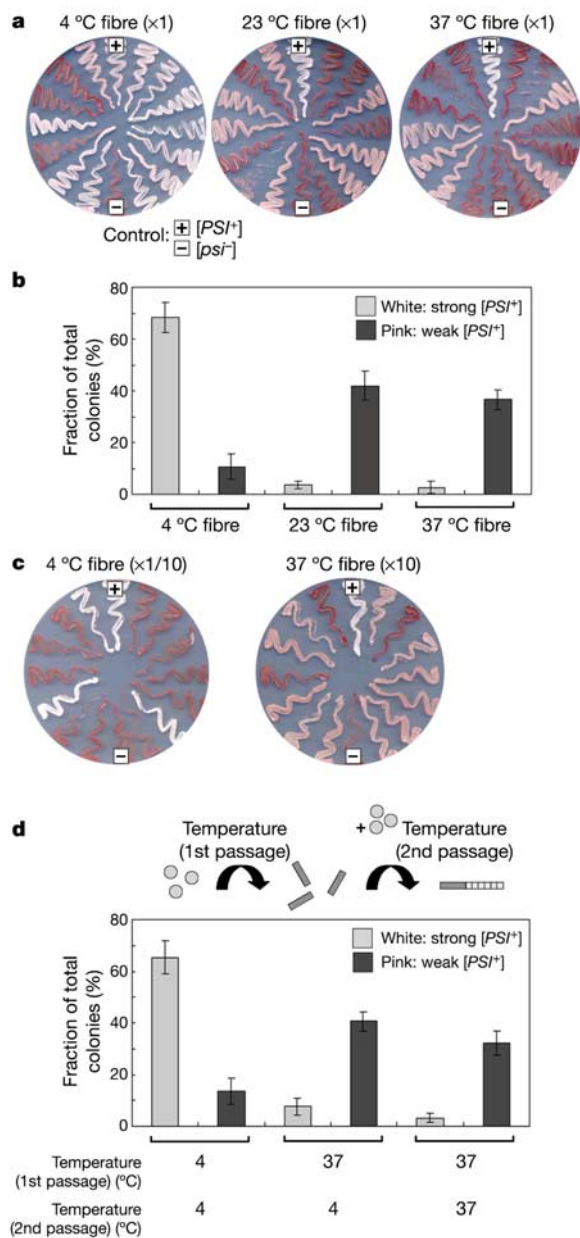


Figure 4 Induction of distinct $[\text{PSI}^+]$ strains by Sup-NM amyloid fibres formed at different temperatures. **a**, Examples of $[\text{PSI}^+]$ strains resulting from infection with 4°C , 23°C and 37°C formed amyloid fibres. White and pink and/or sectorial colonies are strong and weak $[\text{PSI}^+]$ variants, respectively. A substantial majority of the weak $[\text{PSI}^+]$ strains behave like the strains shown in Fig. 2b (that is, pink colour and few visible red sectors). However, the other weak $[\text{PSI}^+]$ strains appear to have an even weaker phenotype leading to a darker pink colour or observable red sectors (for example, Fig. 2a). **b**, Quantification of frequency of $[\text{PSI}^+]$ strains induced by transformation with 4°C , 23°C and 37°C formed amyloid fibres. Preliminary studies indicate that fibres formed by seeding wild-type Sup-NM with the S17R mutant have a low efficiency of infection and lead to weak strains. **c**, Induction of $[\text{PSI}^+]$ using one-tenth the amount of 4°C formed amyloid fibres or tenfold the amount of 37°C formed amyloid fibres used in **a**. **d**, Fraction of strong and weak $[\text{PSI}^+]$ strains induced by the infection of $[\text{psi}^-]$ yeast with amyloids formed by polymerizing Sup-NM at 4°C seeded by a small amount (5%) of 4°C or 37°C fibres, and by Sup-NM fibres formed at 37°C seeded by 37°C fibres.

gates often referred to as amyloids^{1,12,13,22,26,30}. Many amyloid-forming proteins, including those involved in non-infectious diseases, can misfold into more than one form, and these differences can subsequently be propagated by templated conversion⁴. Together with previous studies correlating strains with prion conformation, our work, by directly demonstrating that different amyloid conformations lead to distinct $[PSI^+]$ strains, argues that the ability of proteins to misfold into multiple amyloid forms constitutes the physical foundation of the strain phenomenon.

Note added in proof: We have found that prion particles purified from weak $[PSI^+]$ strains show increased resistance to thermal denaturation in the presence of SDS ($T_m = 74 \pm 6^\circ\text{C}$, $W = 17 \pm 4^\circ\text{C}$) compared with prion particles purified from strong $[PSI^+]$ strains ($T_m = 59 \pm 5^\circ\text{C}$, $W = 27 \pm 5^\circ\text{C}$). This result indicates that conformational differences in infectious prion particles created *in vitro* are faithfully propagated *in vivo*. □

Methods

Construction of plasmids and yeast strains

Plasmid vectors encoding Sup-NM single cysteine mutants for bacterial protein expression were generated using QuikChange mutagenesis (Stratagene). Throughout we used isogenic $[psi^-]$ and $[PSI^+]$ derivatives of 74D-694 [*MATa*, *his3*, *leu2*, *trp1*, *ura3*; suppressible marker *ade1-14(UGA)*]²³. With the exception of Fig. 1c, all strains were $[PIN^+]$. Similar results were obtained with a W303-derived strain. A yeast strain expressing the prion domain from *C. albicans* SUP35 was generated by replacing the wild-type chromosomal locus in the parental 74-D694 strain, as previously described¹⁰.

Preparation of spheroplasts for transformation

Yeast strains were grown in YEPD media to an optical density of 0.5 at 600 nm and successively washed with sterile H₂O, 1 M sorbitol and SCE buffer (1 M sorbitol, 10 mM EDTA, 10 mM dithiothreitol, 100 mM citrate, pH 5.8). Cells were spheroplasted with lyticase (~250 µg for yeast cells of 50 ml YEPD) in SCE buffer at 30 °C for 30 min (see Supplementary Information for a description of lyticase preparation). Spheroplasts were washed with 1 M sorbitol and STC buffer (1 M sorbitol, 10 mM CaCl₂, 10 mM Tris, pH 7.5). Pelleted cells were resuspended in STC buffer and mixed with sonicated amyloid fibres, URA3-marked plasmid (pRS316) (20 µg ml⁻¹) and salmon sperm DNA (100 µg ml⁻¹). Fusion was induced by addition of 9 volumes of PEG buffer (20% (w/v) PEG 8000, 10 mM CaCl₂, 10 mM Tris, pH 7.5) for 30 min. Cells were centrifuged, resuspended in SOS buffer (1 M sorbitol, 7 mM CaCl₂, 0.25% yeast extract, 0.5% bacto-peptone), incubated at 30 °C for 30 min and plated on synthetic media lacking uracil overlaid with top agar (2.5% agar)²⁰. For experiments in Fig. 1b, adenine (20 mg l⁻¹) was included only in the top agar.

Preparation of infectious *in vitro* particles and yeast extracts

Amyloid fibres of Sup35-NM were formed spontaneously at 4 °C, 23 °C or 37 °C or with indicated 5% (w/w) seed by dilution of GuHCl-denatured proteins into 5 mM potassium phosphate buffer (pH 7.4) containing 150 mM NaCl¹⁰. Fibres were collected by centrifugation at 20,000g for 20 min, resuspended with phosphate buffer, and sonicated. The final concentration of amyloid fibres in infection experiments was 2.5 µM unless otherwise indicated. For preparation of yeast extracts, cells were spheroplasted with lyticase and sonicated. Concentration of total protein in the yeast extracts was measured by Bradford assay. For transformations involving extracts, the final concentration of total protein of the crude yeast extract was 200 µg ml⁻¹.

In vivo analysis of prion strains and species specificity

After growth on SD-URA (~7 days), the efficiency of conversion to prion state and phenotypic strength of prion strains was examined by the following colour assay. Positive URA3 transformants were randomly selected and streaked onto YEPD plates containing one-quarter of the standard amount of yeast extract to enhance the colour phenotype. Control strong $[PSI^+]$ and $[psi^-]$ yeast colonies were also streaked onto each plate for comparison. After 3–5 days, colonies were classified as strong $[PSI^+]$ (white), weak (pink) $[PSI^+]$ or $[psi^-]$ (red) strains. In quantification experiments at least three measurements of 56 colonies from independent transformations were made. All quantification experiments used the SD-URA followed by streaking assay rather than direct plating on SD-URA trace ADE.

In vitro analysis of amyloid fibre

Sup-NM proteins C-terminally tagged with 7X-histidine were purified as reported previously²³, except in the case of the cysteine mutants where 1 mM β-mercaptoethanol was included in all purification buffers. The thermal stability of amyloid fibres was examined by SDS-polyacrylamide gel electrophoresis (PAGE) as performed previously¹⁰. The Coomassie-stained band intensities were fitted to a sigmoidal curve with IgorPro3.0 (WaveMetrics), using the following equation: $y = A + B / (1 + \exp((C - x)/D))$, where x , y , C , D indicate temperature, band intensity, melting (T_m) and transition (W) temperatures, respectively. For EPR spectroscopy, single cysteine mutants were labelled by 4-maleimido-2,2,6,6-tetramethyl-1-piperidinoxyl (5 equivalents) (Sigma) in 6 M GuHCl containing 25 mM sodium phosphate buffer (pH 8.0) overnight with ~80% efficiency,

and purified on Ni-NTA agarose (Qiagen). Absence of labelling at non-cysteine residues was confirmed using wild-type Sup-NM. Amyloid fibres were prepared by dilution of the denatured proteins into 5 mM potassium phosphate buffer containing 150 mM NaCl in the presence of 5% (w/w) seed. Fibres were formed at 4 °C and 37 °C with 5% (w/w) seed of spontaneously formed amyloids at 4 °C and 37 °C, respectively. S17R-seeded fibres were formed by polymerization of wild-type Sup-NM at 23 °C with 5% (w/w) seed of spontaneously formed S17R mutant fibres¹⁰. Fibre formation was monitored in parallel by thioflavin T fluorescence¹⁰. After polymerization, fibres were collected by centrifugation at 20,000g for 20 min. Pelleted fibrillar or soluble Sup-NM (~50 µg) was loaded onto a flatcell and EPR spectra were measured in an ER/200D EPR spectrometer (IBM Instruments) at room temperature with 25 mW microwave power and a modulation of 2.0 G at 100 KHz over a scan range of 100 G.

Received 25 November 2003; accepted 5 February 2004; doi:10.1038/nature02392.

- Prusiner, S. B., Scott, M. R., DeArmond, S. J. & Cohen, F. E. Prion protein biology. *Cell* **93**, 337–348 (1998).
- Collinge, J. Prion diseases of humans and animals: their causes and molecular basis. *Annu. Rev. Neurosci.* **24**, 519–550 (2001).
- Aguzzi, A. & Haass, C. Games played by rogue proteins in prion disorders and Alzheimer's disease. *Science* **302**, 814–818 (2003).
- Chien, P., Weissman, J. S. & Depace, A. H. Emerging principles of conformation-based prion inheritance. *Annu. Rev. Biochem.* (in the press).
- Liebmman, S. Progress toward an ultimate proof of the prion hypothesis. *Proc. Natl Acad. Sci. USA* **99**, 9098–9100 (2002).
- Bessen, R. A. *et al.* Non-genetic propagation of strain-specific properties of scrapie prion protein. *Nature* **375**, 698–700 (1995).
- Telling, G. C. *et al.* Evidence for the conformation of the pathologic isoform of the prion protein enciphering and propagating prion diversity. *Science* **274**, 2079–2082 (1996).
- Perez, D. *et al.* A change in the conformation of prions accompanies the emergence of a new prion strain. *Neuron* **34**, 921–932 (2002).
- Chien, P. & Weissman, J. S. Conformational diversity in a yeast prion dictates its seeding specificity. *Nature* **410**, 223–227 (2001).
- Chien, P., DePace, A. H., Collins, S. & Weissman, J. S. Generation of prion transmission barriers by mutational control of amyloid conformations. *Nature* **424**, 948–951 (2003).
- Uptain, S. M. & Lindquist, S. Prions as protein-based genetic elements. *Annu. Rev. Microbiol.* **56**, 703–741 (2002).
- Glover, J. R. *et al.* Self-seeded fibers formed by Sup35, the protein determinant of $[PSI^+]$, a heritable prion-like factor of *S. cerevisiae*. *Cell* **89**, 811–819 (1997).
- King, C. Y. *et al.* Prion-inducing domain 2–114 of yeast Sup35 protein transforms *in vitro* into amyloid-like filaments. *Proc. Natl Acad. Sci. USA* **94**, 6618–6622 (1997).
- Derkatch, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G. & Liebman, S. W. Genesis and variability of $[PSI^+]$ prion factors in *Saccharomyces cerevisiae*. *Genetics* **144**, 1375–1386 (1996).
- Kushnirov, V. V., Kryndushkin, D. S., Boguta, M., Smirnov, V. N. & Ter-Avanesyan, M. D. Chaperones that cure yeast artificial $[PSI^+]$ and their prion-specific effects. *Curr. Biol.* **10**, 1443–1446 (2000).
- Zhou, P. *et al.* The yeast non-Mendelian factor $[ETA^+]$ is a variant of $[PSI^+]$, a prion-like form of release factor eRF3. *EMBO J.* **18**, 1182–1191 (1999).
- Kochneva-Pervukhova, N. V. *et al.* $[PSI^+]$ prion generation in yeast: characterization of the 'strain' difference. *Yeast* **18**, 489–497 (2001).
- Uptain, S. M., Sawicki, G. J., Caughey, B. & Lindquist, S. Strains of $[PSI^+]$ are distinguished by their efficiencies of prion-mediated conformational conversion. *EMBO J.* **20**, 6236–6245 (2001).
- King, C. Y. Supporting the structural basis of prion strains: induction and identification of $[PSI^+]$ variants. *J. Mol. Biol.* **307**, 1247–1260 (2001).
- Sparner, H. E., Santos, A., Szoka, F. C. & Weissman, J. S. Evidence for the prion hypothesis: Induction of the yeast $[PSI^+]$ factor by *in vitro*-converted sup35 protein. *Science* **289**, 595–599 (2000).
- Derkatch, I. L., Bradley, M. E., Zhou, P., Chernoff, Y. O. & Liebman, S. W. Genetic and environmental factors affecting the de novo appearance of the $[PSI^+]$ prion in *Saccharomyces cerevisiae*. *Genetics* **147**, 507–519 (1997).
- Maddelaine, M.-L., Dos Reis, S., Duvezin-Caubet, S., Couлары-Salin, B. & Saupe, S. J. Amyloid aggregates of the HET-s prion protein are infectious. *Proc. Natl Acad. Sci. USA* **99**, 7402–7407 (2002).
- Santos, A., Chien, P., Osheroch, L. Z. & Weissman, J. S. Molecular basis of a yeast prion species barrier. *Cell* **100**, 277–288 (2000).
- Chernoff, Y. O. *et al.* Evolutionary conservation of prion-forming abilities of the yeast Sup35 protein. *Mol. Microbiol.* **35**, 865–876 (2000).
- Kushnirov, V. V., Kochneva-Pervukhova, N. V., Chechenova, M. B., Frolova, N. S. & Ter-Avanesyan, M. D. Prion properties of the Sup35 protein of yeast *Pichia methanolica*. *EMBO J.* **19**, 324–331 (2000).
- Baxa, U., Speransky, V., Steven, A. C. & Wickner, R. B. Mechanism of inactivation on prion conversion of the *Saccharomyces cerevisiae* Ure2 protein. *Proc. Natl Acad. Sci. USA* **99**, 5253–5260 (2002).
- DePace, A. H. & Weissman, J. S. Origins and kinetic consequences of diversity in Sup35 yeast prion fibers. *Nature Struct. Biol.* **9**, 389–396 (2002).
- Rice, S. *et al.* Thermodynamic properties of the kinesin neck-region docking to the catalytic core. *Biophys. J.* **84**, 1844–1854 (2003).
- McHaourab, H. S., Lietzow, M. A., Hideg, K. & Hubbell, W. L. Motion of spin-labeled side chains in T4 lysozyme: Correlation with protein structure and dynamics. *Biochemistry* **35**, 7692–7704 (1996).
- Sondheimer, N. & Lindquist, S. L. Rnq1: an epigenetic modifier of protein function in yeast. *Mol. Cell* **5**, 163–172 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Collins, J. Newman, L. Osherovich, K. Tipton and members of the Weissman laboratory for discussion and reading of the manuscript. M.T. was supported by JSPS postdoctoral fellowships for research abroad. P.C. was supported by National Science Foundation Graduate Fellowships and the ARCS foundation. Funding was also provided by Howard Hughes Medical Institute, The David and Lucile Packard Foundation and the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S.W. (jsw1@itsa.ucsf.edu).

The candidate tumour suppressor protein ING4 regulates brain tumour growth and angiogenesis

Igor Garkavtsev¹, Sergey V. Kozin¹, Olga Chernova², Lei Xu¹, Frank Winkler¹, Edward Brown¹, Gene H. Barnett² & Rakesh K. Jain¹

¹Edwin L. Steele Laboratory for Tumour Biology, Department of Radiation Oncology Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA

²The Brain Tumour Institute, The Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA

Gliomas are the most common primary tumours of the central nervous system, with nearly 15,000 diagnosed annually in the United States and a lethality approaching 80% within the first year of glioblastoma diagnosis¹. The marked induction of angiogenesis in glioblastomas suggests that it is a necessary part of malignant progression²; however, the precise molecular mechanisms underlying the regulation of brain tumour growth and angiogenesis remain unresolved. Here we report that a candidate tumour suppressor gene, *ING4*, is involved in regulating brain tumour growth and angiogenesis. Expression of *ING4* is significantly reduced in gliomas as compared with normal human brain tissue, and the extent of reduction correlates with the progression from lower to higher grades of tumours. In mice, xenografts of human glioblastoma U87MG, which has decreased expression of *ING4*, grow significantly faster and have higher vascular volume fractions than control tumours. We show that *ING4* physically interacts with p65 (RelA) subunit of nuclear factor NF- κ B, and that *ING4* regulates brain tumour angiogenesis through transcriptional repression of NF- κ B-responsive genes. These results indicate that *ING4* has an important role in brain tumour pathogenesis.

Tumour angiogenesis is significantly increased in malignant gliomas². The angiogenic stimulators and inhibitors comprise growth factors, proteases, tumour suppressors, oncogenes and cytokines^{3,4}. One of the key tumour suppressor proteins, p53, is a strong suppressor of angiogenesis in several cancers, including astrocytic tumours^{5,6}. ING family proteins and p53 form a transcriptional complex that is required for the activation of p53-responsive genes and that mediates growth arrest, replicative senescence, apoptosis and DNA repair^{7–13}. We searched GenBank for sequences with homology to *ING1* to identify additional members of this family that might be involved in tumour growth and angiogenesis. The *ING4* protein has nuclear localization signals (three blocks of the positively charged amino acid lysine) in the 127–164 amino acid regions (Supplementary Fig. S1a). We confirmed its nuclear localization by indirect immunofluorescence (Supplementary Fig. S1c). Expression of the *ING4* gene was detected in all human tissues examined (Supplementary Fig. S1b).

Because *ING1* expression is repressed in various cancers^{14–17}, we anticipated that expression of *ING4* would also be reduced in some tumours. Gliomas are divided into grade II (low grade), grade III (anaplastic) and grade IV, which includes the most malignant neoplasms with an astrocytic phenotype, glioblastoma multiforme (GBM)^{18,19}. To test whether the *ING4* gene is associated with brain tumour growth and progression, we analysed its expression in 50 glioma tumours of different grades of malignancy by using real-time polymerase chain reaction with reverse transcription (RT-PCR; Fig. 1a). Average levels of *ING4* messenger RNA were 2–3 times lower in low-grade gliomas (grade II–III), and six times lower in glioblastomas (grade IV) than in normal brain tissue. These results indicate that expression of *ING4* is significantly suppressed in brain tumours and that this loss of expression correlates with tumour grades.

To confirm downregulation of the *ING4* gene in brain tumours and to study heterogeneity in the expression of *ING4*, we carried out immunohistochemical analysis. Immunostaining of normal brain tissue located adjacent to glioblastoma with antibodies against

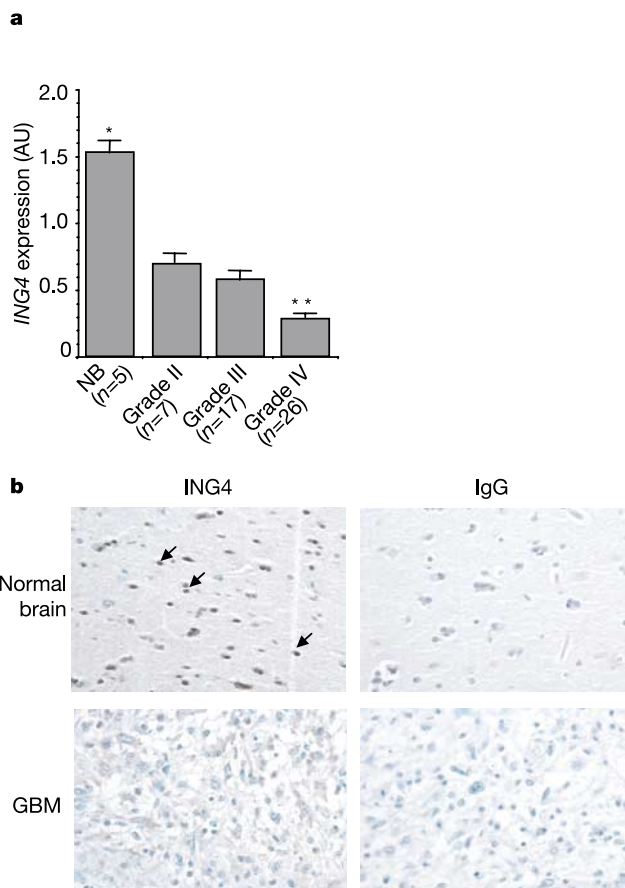


Figure 1 Expression of *ING4* is downregulated in human gliomas and is negatively correlated with their progression. **a**, Real-time RT-PCR was used to quantify *ING4* expression. Brain tumours were grouped according to the grade of malignancy: grade II, oligodendroglioma ($n = 5$) and astrocytoma ($n = 2$); grade III, anaplastic oligodendroglioma ($n = 11$), anaplastic astrocytoma ($n = 3$) and anaplastic oligoastrocytoma ($n = 3$); grade IV, GBM ($n = 26$). RNA derived from several human normal brain regions (NB) was used as a control ($n = 5$). Columns and bars show the mean and s.e.m., respectively. The difference is significant between the control and all tumour groups ($*P < 0.05$), and between lower-grade gliomas (grades II and III) and GBM ($**P < 0.05$). **b**, Immunohistochemical analysis of *ING4* protein in brain tumours. Shown are representative fields of normal brain adjacent to the tumour and the tumour area stained with anti-*ING4* and nonspecific IgG polyclonal antibodies. Arrows indicate *ING4*-specific nuclear staining.

ING4 protein showed a strong nuclear signal, which confirmed our *in vitro* observations regarding the nuclear localization of *ING4* (Fig. 1b). The adjacent area of the same specimen containing glioblastoma showed significantly less nuclear staining in a highly cellular tumour tissue, confirming our RT-PCR data.

As extensive microvascular proliferation is a hallmark of GBM, we speculated that alterations in *ING4* expression might affect this process. To test this hypothesis, we grafted human glioblastoma U87MG cells into the cranial window, which facilitates *in vivo* real-time imaging of brain tumour angiogenesis and growth using intravital microscopy (IVM)²⁰. We generated derivatives of the U87MG glioblastoma cell line with increased or reduced expression of *ING4* by transfecting the cells with plasmids carrying the *ING4* gene in sense and antisense orientations, respectively. Overexpression and downregulation of *ING4* protein in the respective sense and antisense cells were confirmed by western blot analysis (Fig. 2a).

Source tumours derived from these cell lines were implanted in the cranial windows (Fig. 2b). Antisense-*ING4* (as-*ING4*) tumours grew faster than both control and *ING4*-overexpressing tumours ($P < 0.05$). Most as-*ING4* and control tumours filled the whole space under the cranial window on day 12, precluding further quantification of tumour area. The slowest rate of tumour growth was observed in tumours overexpressing *ING4* ($P < 0.05$). There was a direct correlation between thickness of the tumours and size of their surface area (Fig. 2c). IVM showed that as-*ING4* tumours were more robustly vascularized, with a higher vessel density and a higher frequency of haemorrhage than control tumours (25% versus 0%).

To characterize the differences in tumour vasculature further, we generated images of the *ING4*-modified and control U87MG tumours by using two-photon IVM (Fig. 2d–f). Compared with control tumours (Fig. 2d, g), U87MG implants with downregulated *ING4* (as-*ING4*) were more angiogenic, as characterized by an increase in the vascular volume fraction ($P < 0.05$; Fig. 2e, g). Conversely, overexpression of the *ING4* gene resulted in a less-developed vascular network with a significant decrease in vascular

volume fraction relative to controls ($P < 0.05$; Fig. 2f, g). Furthermore, analysis of mean vessel diameter and vessel density from higher magnification images showed that there was a significant decrease ($P < 0.005$) in both parameters in *ING4*-overexpressing tumours, as compared with controls and as-*ING4* tumours. By contrast, as-*ING4* tumours showed an increase in the proportion of very large vessels, as compared with the other two groups (Supplementary Fig. S2). Together, these findings indicate that *ING4* negatively affects angiogenesis during tumour development and consequently suppresses tumour growth rate.

To determine the molecular mechanisms of *ING4*-mediated anti-angiogenic regulation, we analysed expression profiles in *ING4*-modified U87MG cells by using specialized complementary DNA microarrays containing 288 angiogenesis-related genes. The expression of six genes, including two interleukins (IL-6 and IL-8), prostaglandin-endoperoxide synthase 2 (Cox-2) and colony-stimulating factor 3, was increased in as-*ING4* cells. Notably, these genes are regulated by the transcription factor NF- κ B and frequently show abnormal expression in cancer^{21,22}. These expression changes were confirmed by northern blot analysis (Fig. 3a). These data indicate that suppression of *ING4* correlates with an increase in expression of NF- κ B-responsive genes.

To confirm that alterations in IL-8 expression in our *in vivo* model correspond to *ING4* modulation in the implanted U87MG cells, we used enzyme-linked immunosorbent assay (ELISA) to measure IL-8 protein in tumour tissues excised from the cranial window (Fig. 3b). IL-8 increased significantly from 80 to 780 pg mg⁻¹ in response to the downregulation of *ING4*. IL-8 induces endothelial cell chemotactic and proliferative activities and mediates neovascularization in tumours^{23–25}. We found that the *ING4* gene is not involved directly in regulating vascular endothelial growth factor (VEGF) *in vitro* or *in vivo* (Fig. 3a, c).

NF- κ B is involved in regulating the adaptive immune responses, cell survival, angiogenesis, tumorigenesis and metastasis^{21,22}. The ability of *ING4* and NF- κ B to interact was tested in co-immuno-

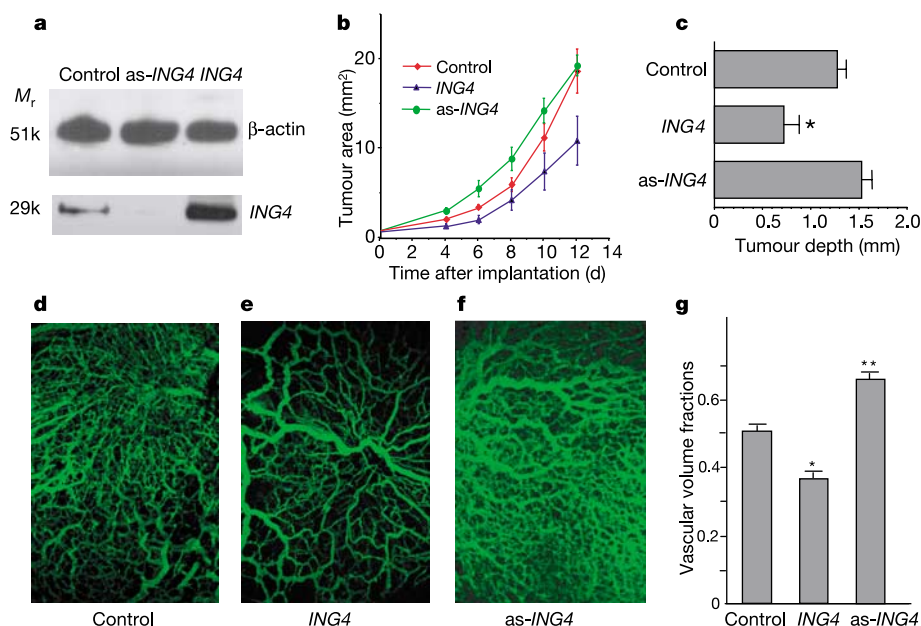


Figure 2 *ING4* regulates tumour growth and angiogenesis *in vivo*. **a**, Level of *ING4* protein in modified U87MG cells. A total of 20 μ g of protein per line was analysed by western blotting using an anti-*ING4* antibody. β -actin was used as an equal loading control. **b**, Growth dynamics of human U87MG cell lines expressing control vector, and antisense and sense *ING4* in the cranial windows of SCID mice. **c**, Thickness of the tumours determined at autopsy on day 13 (* $P < 0.05$). **d–f**, Images of angiogenic vasculature in U87MG tumour variants with different *ING4* status: control vector expression (**d**), *ING4*

overexpression (**e**) and *ING4* downregulation (**f**). To obtain high-resolution images of tumour vasculature, a fluorescent contrast agent (FITC-dextran) was injected into the tail veins of mice, and the vessels were observed by two-photon IVM. Image height, 2 mm. **g**, Analysis of the vascular volume fractions in U87MG-modified tumours. Ten mice were used per group. Compared with controls, as-*ING4* tumours had higher (** $P < 0.05$) and *ING4* tumours had lower (* $P < 0.05$) vascular volume fraction. In **b**, **c** and **g**, data values and bars show the mean and s.e.m., respectively.

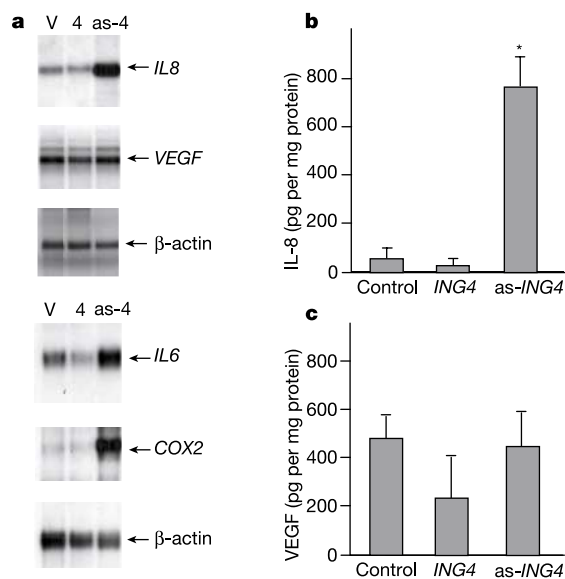


Figure 3 Downregulation of the *ING4* gene activates the angiogenic *IL8* gene *in vitro* and *in vivo*. **a**, *ING4*-dependent expression of the *IL8*, *IL6* and *COX2* genes. Expression of these genes was detected in U87MG cell lines with different *ING4* status: control vector expression (V), *ING4* overexpression (4) and *ING4* downregulation (as-4). mRNA was isolated from U87MG cells and analysed by northern blotting for expression of the *IL8*, *IL6*, *COX2*, *VEGF* and β -actin genes. **b**, **c**, Detection of IL-8 (**b**) and VEGF (**c**) protein in mouse brain tumours. Expression of IL-8 and VEGF in tumours was determined by ELISA. Three mice were used for each group. * $P < 0.001$, compared with control and *ING4*.

precipitation experiments. Flag-*ING4* and control Flag constructs were transiently transfected into cells, and antibodies against Flag were used to precipitate Flag-containing protein complexes from cellular extracts. The presence of NF- κ B protein in precipitates was monitored by immunoblotting with polyclonal antibodies against the p65 (RelA) subunit of NF- κ B. p65 was co-immunoprecipitated with *ING4*, but not with *ING4*-pre-incubated Flag-blocking peptide, showing that there is a physical association between the large subunit of NF- κ B and *ING4* (Fig. 4a).

To confirm this interaction, mammalian two-hybrid experiments were carried out (Fig. 4b). A pair of GAL4-*ING4* and VP16-p65 fusion proteins induced activation of the reporter construct 4–5 times more than did the negative controls, verifying the physical association between *ING4* and p65. To confirm that endogenous *ING4* and NF- κ B form a stable complex, co-immunoprecipitation assays were done with antibodies against p65, and immunoprecipitated proteins were detected by western blot using antibodies against *ING4* protein. The assay showed that *ING4* physically interacts with NF- κ B (Fig. 4c). Specific immunoprecipitation was not observed with normal rabbit immunoglobulin- γ (IgG), which was used as a negative control.

To evaluate whether *ING4* regulates NF- κ B activity, we carried out a gel mobility shift assay. U87MG cells have constitutively activated NF- κ B. Biotin-labelled NF- κ B was incubated with nuclear protein extracts from derivative U87MG cells with modified expression of the *ING4* gene. Downregulation of *ING4* resulted in a substantial increase in the binding activity of NF- κ B (Fig. 4d). The binding of this protein to the probe was sequence-specific, as it was blocked by competition with an unlabelled NF- κ B oligomer but not by an unrelated Oct-1 oligomer. To confirm that *ING4* binds to p65, we carried out a supershift assay using antibodies against p65

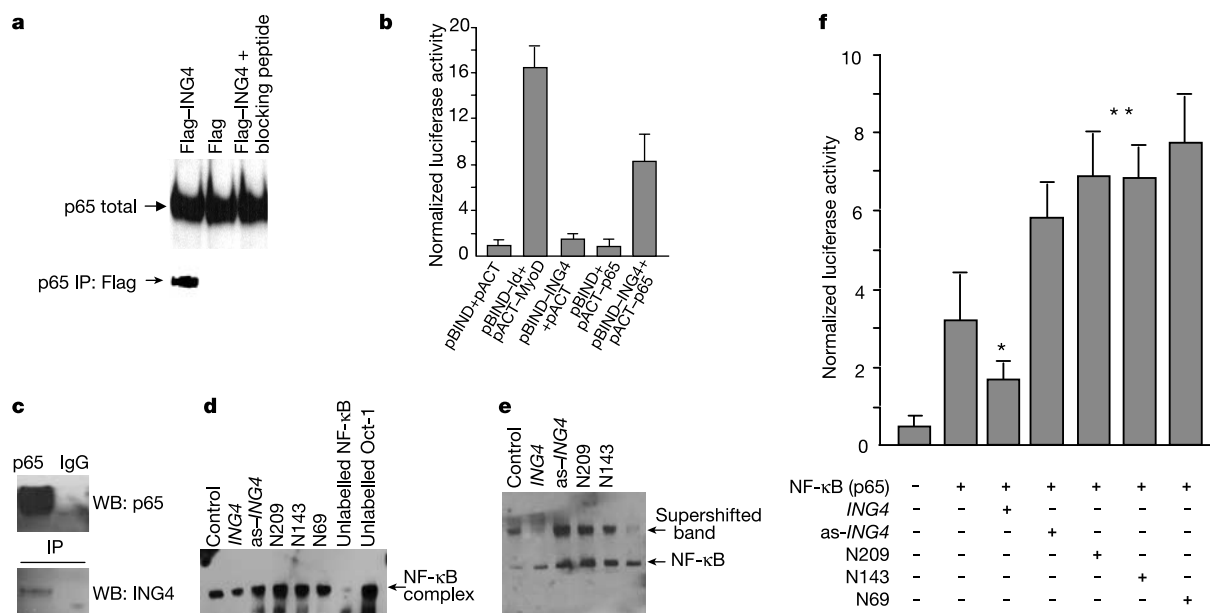


Figure 4 Interaction of *ING4* and NF- κ B decreases NF- κ B binding activity and inhibits transactivation of NF- κ B-responsive genes. **a**, Co-immunoprecipitation of *ING4* and the p65 subunit of NF- κ B (RelA). Cells were transfected with vectors expressing control Flag or Flag-*ING4*. Total cell lysates and immunoprecipitates were analysed by immunoblotting with antibodies against p65. **b**, Interaction between *ING4* and p65 assessed by a mammalian two-hybrid system. The indicated combinations of plasmid (Methods) were transfected together with the reporter plasmid pG5-luc into U87MG cells, and luciferase activity was assayed at 48 h. **c**, Gel mobility shift assay of nuclear proteins extracted from *ING4*-modified U87MG stable cell lines. All lanes contained the labelled NF- κ B probe. Arrow indicates the NF- κ B complex. **d**, Supershift assay. Anti-p65 was

incubated with the gel-shift binding reaction. All lanes contained the labelled NF- κ B probe. Arrow indicates bands supershifted by antibodies against p65. **e**, U87MG cells were transiently transfected with an NF- κ B-dependent luciferase reporter plasmid (5X κ B-luc) containing NF- κ B-binding elements together with NF- κ B and the indicated *ING4* constructs. Cell lysates were collected 48 h after transfection. The luciferase activities were normalized for transfection efficiency (by *Renilla* luciferase activity). Results are the mean $\pm$ s.d. of five independent experiments. Differences in the relative luciferase activities were analysed by Student's *t*-test. * $P < 0.05$, *ING4* compared with control; ** $P < 0.05$, as-*ING4*, *ING4*(N209), *ING4*(N143) and *ING4*(N69) compared with control.

(Fig. 4e). This assay showed that ING4 binds to p65 and is a negative regulator of NF- κ B activity.

To test directly the ability of *ING4* to modulate expression of the NF- κ B target genes, U87MG cells were transfected with an integrated NF- κ B-dependent luciferase reporter plasmid (5X κ B-luc) containing tandem repeats of NF- κ B-binding elements and different combinations of the sense or antisense *ING4* constructs. An as-*ING4* cDNA stimulated NF- κ B-mediated transcription, as judged by a 1.5–2-fold increase in luciferase activity as compared with U87MG cells transfected with an NF- κ B-expressing plasmid alone. The sense *ING4* construct had the opposite effect. Vectors expressing the truncated versions (N209, N143 and N69) of *ING4* had an effect similar to that of the as-*ING4* construct, presumably through a dominant-negative activity. These data indicate that expression of full-length *ING4* cDNA is consistent with blockage

of NF- κ B activation, whereas expression of *ING4* fragments and the as-*ING4* construct has a stimulatory effect on the NF- κ B-dependent transactivation of responsive genes.

To prove that *ING4*-mediated IL-8 regulation has a significant influence on brain tumour growth and angiogenesis, short interfering RNA targeted at IL-8 (*siIL8*) was used to inhibit IL-8 synthesis. Downregulation of IL-8 in U87MG as-*ING4* cells was confirmed by ELISA assay (Fig. 5a). Both types of cell grew at a similar rate *in vitro*. The U87MG as-*ING4* *siIL8* and U87MG as-*ING4* stable cell lines were used as tumour sources. On average, U87MG as-*ING4* *siIL8* tumours grew significantly slower than their U87MG as-*ING4* counterparts, and in almost half of the samples they did not grow at all. Figure 5b shows the median growth curves in both groups. The U87MG as-*ING4* *siIL8* tumours showed a decrease in angiogenesis comparable to that of the *ING4* tumours described above (Figs 5d and 2d). Together, these findings indicate that the regulation of IL-8 by *ING4* affects tumour angiogenesis and directly influences brain tumour growth.

Our results collectively show that *ING4* physically interacts with NF- κ B, forming a transcriptional complex, and represses NF- κ B-responsive genes (Fig. 5e). The direct participation of *ING4* in the NF- κ B signalling pathway indicates that *ING4* is a candidate tumour suppressor gene and that its inactivation contributes to brain tumour growth and angiogenesis. □

Methods

Plasmids, antibodies, cells and ELISA

The full-length *ING4* cDNA was subcloned to a pcDNA3.1 plasmid under the control of the cytomegalovirus (CMV) promoter in both sense and antisense orientations. Truncated *ING4* fragments encoding 209 (N209), 143 (N143) and 69 (N69) amino acids were also cloned into pcDNA3.1. We transfected U87MG cells (7.5×10^5) with 2 μ g of plasmid by using Lipofectamine (Invitrogen) in accordance with the manufacturer's instructions. The polyclonal rabbit antibodies were raised against an NH₂-EKQIESDYSSSSKGGK-OH peptide of *ING4*. All cell lines were maintained in DMEM medium with 10% fetal bovine serum. We measured IL-8 and VEGF protein expression by ELISA using a Quantikine kit (R&D Systems).

Intravital microscopy

Human U87MG cell lines expressing the indicated *ING4* constructs were initially transplanted subcutaneously into severe combined immunodeficient (SCID) mice to generate source tumours. These tumours were excised, cut into small chunks (diameter 0.6 ± 0.1 mm) and implanted on the cerebrum into cranial windows prepared in SCID mice by standard procedures²⁰. We monitored tumour growth every other day by IVM. Tumour vessels were highlighted by an intravenous injection of 0.1 ml of 10 mg ml^{-1} fluorescein isothiocyanate (FITC)-conjugated dextran (relative molecular mass 2,000,000). We generated image stacks consisting of 10 images spaced 20- μ m apart using a 5X/0.15 NA lens, or 20 images spaced 5- μ m apart using a 20X/0.4 NA lens. To calculate the two-dimensional volume fraction, maximum intensity projections (MIPs) of each 5X stack were created, all images were subjected to a cut-off with a common threshold, and the total number of pixels above threshold was calculated and normalized by the total image pixel number. To calculate average vessel diameter and number, we created MIPs of 20X stacks and superimposed 4X4 grids. Wherever grid lines intersected vessels, the vessel diameter was measured.

Real-time PCR

Human glioblastoma tissue specimens were collected in accordance with IRB approval from the Cleveland Clinic Foundation. Total RNA from frozen tumour specimens was extracted by using TRIzol reagent (Invitrogen) in accordance with the manufacturer's protocol. Taqman RT-PCR was conducted using the ABI PRISM 7700 Sequence Detection System (Applied Biosystems), PCR was done in triplicate for each sample, and all experiments were repeated three times. Human B2-microglobulin (B2M) was used as a control in the PCR reactions. We used the following set of primers and TaqMan probes (Applied Biosystems) for real-time PCR: for *ING4*, forward and reverse primers (5' to 3', CCA GGT CTC CTA TGG AGA GAT GA and CCC GAG GCT TGG TTG TCA) and TaqMan probe (6FAM ACC CTG ATT GTT CCA TTG AGT GGT TCC A TAMRA); for B2M, forward and reverse primers (5' to 3', TGA GTA TGC CTG CCG TGT GA and CAA ACC TCC ATG ATG CTG CTT) and TaqMan probe (6FAM TGT CTC GAT CCC ACT TAA CTA TCT TGG GCT GT TAMRA).

RNA interference

To obtain a stable short interfering RNA (siRNA) construct for silencing the *IL8* gene, we used primers 5'-GATCCCTGCCAAGGAGTGCTAAAGATTCAGAGATCTTTAGCAC TCCTTGGCATTTTGTGAAA-3' and 5'-AGCTTTTCCAAAAATGCCAAGGAGTGCTAAAGATCTCTTGAA TCTTTTAGCACTCTTGGCAGG-3' and subcloned the product into a pSilencer 3.1-H1 hygro plasmid (Ambion).

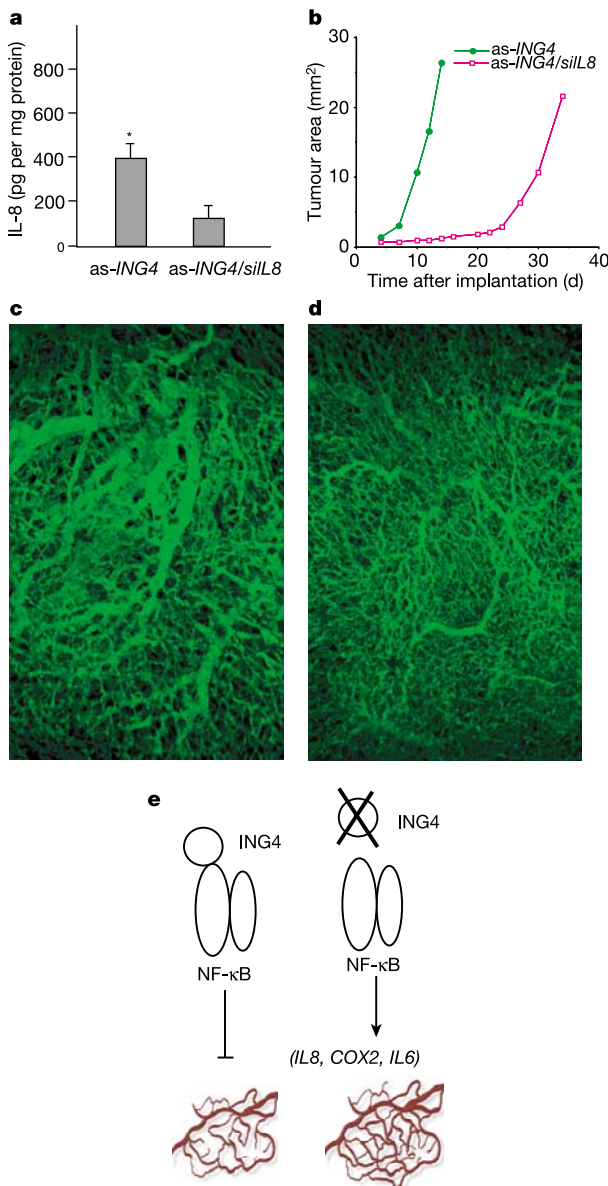


Figure 5 *ING4*-mediated regulation of IL-8 significantly influences brain tumour growth and angiogenesis. **a**, Detection of IL-8 in cell-culture supernatants by ELISA. * $P < 0.05$, as-*ING4* compared with as-*ING4* plus *siIL8*. **b**, Growth of median-size tumours in groups ($n = 7$) of U87MG as-*ING4* tumours and U87MG as-*ING4* *siIL8* tumours, in which expression of IL-8 was downregulated as a control. **c**, **d**, Two-photon imaging of angiogenesis in the U87MG as-*ING4* (**c**) and U87MG as-*ING4* *siIL8* (**d**) tumour variants. **e**, Model of the consequences of *ING4* loss on NF- κ B activity.

Immunoprecipitation and two-hybrid assays

Transfected cells were washed with PBS and treated with RIPA buffer with no SDS. Mouse monoclonal antibody against Flag (M2, Sigma) was added to cell extract and incubated overnight at 4 °C. Protein-A-conjugated Sepharose equilibrated in RIPA was then added and incubated for 2 h. The beads were extensively washed with ice-cold RIPA and the precipitate was dissolved in a sample buffer. Polyclonal antibody against p65 (Rockland) was used for western blot analysis. For mammalian two-hybrid experiments, the *ING4* gene was subcloned in-frame into pBIND and fused with the yeast GAL4 DNA-binding domain. p65 was subcloned in-frame into pACT and fused to the VP16 activation domain of herpes simplex virus. A series of constructs was used as controls: pBIND–Id and pACT–MyoD, containing GAL4–Id and VP16–MyoD fusion proteins, respectively, were used as positive controls; and pBIND and pACT, pBIND–ING4 and pACT, and pBIND and pACT–p65 were used as negative controls. Protein–protein interaction was determined by the activation of firefly luciferase.

Reporter gene assay

Cells (1×10^5 per well) were transfected with plasmids containing NF- κ B-binding elements by Lipofectamine (Invitrogen). Two days after transfection, cells were lysed in Reporter Lysis Buffer (Promega), and luciferase activity was measured by a luminometer using an Enhanced Luciferase Assay kit (BD Biosciences).

Electrophoretic mobility shift assay

We carried out electrophoretic mobility shift assays (EMSAs) on nuclear extracts prepared as described²⁵ using the following double-stranded oligonucleotides: NF- κ B, 5'-AGT TGA GGG GAC TTT CCC AGG C-3'; and nonspecific oligonucleotide (Oct-1), 5'-TGT CGA ATG CAA ATC ACT AGA A-3'. The oligonucleotide was 5'-end-labelled with biotin and annealed by standard procedures. Binding reaction was carried out by pre-incubating nuclear extract protein (5 μ g) in 2.5% glycerol, 1 μ g of poly(dI-dC), 50 mM KCl and 5 mM MgCl₂ at room temperature for 15 min. For competition assays, we added a tenfold molar excess of unlabelled oligonucleotide to the binding reaction. Samples were loaded on a 5% polyacrylamide gel. Electrophoresis was done at room temperature for 3 h at 100 V. The gel was then transferred to a nylon membrane and the biotin end-labelled DNA was detected using streptavidin-conjugated horseradish peroxidase. For supershift assay, antibody against p65 (Santa Cruz Biotech) was added to the binding reaction for 45 min on ice.

Received 16 October 2003; accepted 9 January 2004; doi:10.1038/nature02329.

1. Maher, E. A. *et al.* Malignant glioma: genetics and biology of a grave matter. *Genes Dev.* **15**, 1311–1333 (2001).
2. Plate, K. H. & Risau, W. Angiogenesis in malignant gliomas. *Glia* **15**, 339–347 (1995).
3. Kerbel, R. & Folkman, J. Clinical translation of angiogenesis inhibitors. *Nature Rev. Cancer* **2**, 727–739 (2002).
4. Carmeliet, P. & Jain, R. K. Angiogenesis in cancer and other diseases. *Nature* **407**, 249–257 (2000).
5. Van Meir, E. G. *et al.* Release of an inhibitor of angiogenesis upon induction of wild type p53 expression in glioblastoma cells. *Nature Genet.* **8**, 171–176 (1994).
6. Dameron, K. M., Volpert, O. V., Tainsky, M. A. & Bouck, N. Control of angiogenesis in fibroblasts by p53 regulation of thrombospondin-1. *Science* **265**, 1582–1584 (1994).
7. Garkavtsev, I., Kazarov, A., Gudkov, A. & Riabowol, K. Suppression of the novel growth inhibitor p33^{ING1} promotes neoplastic transformation. *Nature Genet.* **14**, 415–420 (1996).
8. Garkavtsev, I. & Riabowol, K. Extension of the replicative lifespan of human diploid fibroblasts by inhibition of the p33^{ING1} candidate tumour suppressor. *Mol. Cell. Biol.* **17**, 2014–2019 (1997).
9. Garkavtsev, I. *et al.* The candidate tumour suppressor p33^{ING1} cooperates with p53 in cell growth control. *Nature* **391**, 295–298 (1998).
10. Shiseki, M. *et al.* p29^{ING4} and p28^{ING5} bind to p53 and p300, and enhance p53 activity. *Cancer Res.* **63**, 2373–2378 (2003).
11. Skowrya, D. *et al.* Differential association of products of alternative transcripts of the candidate tumour suppressor ING1 with the mSin3/HDAC1 transcriptional corepressor complex. *J. Biol. Chem.* **276**, 8734–8739 (2001).
12. Nourani, A. *et al.* Role of an ING1 growth regulator in transcriptional activation and targeted histone acetylation by the NuA4 complex. *Mol. Cell. Biol.* **21**, 7629–7640 (2001).
13. Leung, K. M. *et al.* The candidate tumour suppressor ING1b can stabilize p53 by disrupting the regulation of p53 by MDM2. *Cancer Res.* **62**, 4890–4893 (2002).
14. Toyama, T. *et al.* Suppression of ING1 expression in sporadic breast cancer. *Oncogene* **18**, 5187–5193 (1999).
15. Oki, E., Maehara, Y., Tokunaga, E., Kakeji, Y. & Sugimachi, K. Reduced expression of p33^{ING1} and the relationship with p53 expression in human gastric cancer. *Cancer Lett.* **147**, 157–162 (1999).
16. Chen, B., Campos, E. L., Crawford, R., Martinka, M. & Li, G. Analyses of the tumour suppressor ING1 expression and gene mutation in human basal cell carcinoma. *Int. J. Oncol.* **22**, 927–931 (2003).
17. Gunduz, M. *et al.* Genomic structure of the human ING1 gene and tumour-specific mutations detected in head and neck squamous cell carcinomas. *Cancer Res.* **60**, 3143–3146 (2000).
18. Brat, D. J., Castellano-Sanchez, A., Kaur, B. & Van Meir, E. G. Genetic and biologic progression in astrocytomas and their relation to angiogenic dysregulation. *Adv. Anat. Pathol.* **9**, 24–36 (2002).
19. Kleihues, P. *et al.* The WHO classification of tumours of the nervous system. *J. Neuropathol. Exp. Neurol.* **61**, 215–225 (2002).
20. Jain, R. K., Munn, L. L. & Fukumura, D. Dissecting tumour pathophysiology using intravital microscopy. *Nature Rev. Cancer* **2**, 266–276 (2002).
21. Ghosh, S. & Karin, M. Missing pieces in the NF- κ B puzzle. *Cell* **109** (suppl.), S81–S96 (2002).
22. Karin, M., Cao, Y., Greten, F. R. & Li, Z. W. NF- κ B in cancer: from innocent bystander to major culprit. *Nature Rev. Cancer* **2**, 301–310 (2002).
23. Desbaillets, I., Diserens, A. C., de Tribolet, N., Hamou, M. F. & Van Meir, E. G. Regulation of interleukin-8 expression by reduced oxygen pressure in human glioblastoma. *Oncogene* **18**, 1447–1456 (1999).
24. Desbaillets, I., Diserens, A. C., Tribolet, N., Hamou, M. F. & Van Meir, E. G. Upregulation of

- interleukin 8 by oxygen-deprived cells in glioblastoma suggests a role in leukocyte activation, chemotaxis, and angiogenesis. *J. Exp. Med.* **186**, 1201–1212 (1997).
25. Xu, L., Xie, K., Mukaida, N., Matsushima, K. & Fidler, I. J. Hypoxia-induced elevation in interleukin-8 expression by human ovarian carcinoma cells. *Cancer Res.* **59**, 5822–5829 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. di Tomaso, P. Vitello, S. Roberge and A. Merkulova for technical assistance; and B. Seed, D. Fukamura and D. Duda for comments on the manuscript. This work was supported by grants from the National Cancer Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.G. (igorg@stele.mgh.harvard.edu).

Survival signalling by Akt and eIF4E in oncogenesis and cancer therapy

Hans-Guido Wendel¹, Elisa de Stanchina¹, Jordan S. Fridman^{1*}, Abba Malina², Sagarika Ray¹, Scott Kogan³, Carlos Cordon-Cardo⁴, Jerry Pelletier² & Scott W. Lowe¹

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

²McGill University, Montreal, H3G1Y6, Quebec, Canada

³University of California-San Francisco, San Francisco, California 94143-0128, USA

⁴Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

* Present address: Incyte Corporation, Wilmington, Delaware 19880, USA

Evading apoptosis is considered to be a hallmark of cancer, because mutations in apoptotic regulators invariably accompany tumorigenesis¹. Many chemotherapeutic agents induce apoptosis, and so disruption of apoptosis during tumour evolution can promote drug resistance². For example, Akt is an apoptotic regulator that is activated in many cancers and may promote drug resistance *in vitro*³. Nevertheless, how Akt disables apoptosis and its contribution to clinical drug resistance are unclear. Using a murine lymphoma model, we show that Akt promotes tumorigenesis and drug resistance by disrupting apoptosis, and that disruption of Akt signalling using the mTOR inhibitor rapamycin reverses chemoresistance in lymphomas expressing Akt, but not in those with other apoptotic defects. eIF4E, a translational regulator that acts downstream of Akt and mTOR, recapitulates Akt's action in tumorigenesis and drug resistance, but is unable to confer sensitivity to rapamycin and chemotherapy. These results establish Akt signalling through mTOR and eIF4E as an important mechanism of oncogenesis and drug resistance *in vivo*, and reveal how targeting apoptotic programmes can restore drug sensitivity in a genotype-dependent manner.

Apoptosis is controlled by a complex network of proliferation and survival genes that is frequently disrupted during tumour evolution. For example, the phosphatidylinositol-3-OH kinase (PI(3)K) pathway integrates receptor tyrosine kinase signalling with the apoptotic network^{4,5}. One mediator of PI(3)K signalling is the Akt/protein kinase B (PKB) kinase, which phosphorylates multiple downstream effectors that ultimately produce global changes in cellular physiology. How Akt promotes survival is controversial, but it may involve direct phosphorylation of apoptotic regulators, increased cell cycle progression, decreased transcription of pro-apoptotic genes through inhibition of forkhead transcription factors, altered metabolism, or changes in the translation of messenger RNAs that ultimately control cell death⁵.

Mutations that activate the PI(3)K–Akt pathway, including amplifications of PI(3)K pathway components and inactivation of the negative regulator PTEN, are common in human malignancies^{6–8}. Therefore, drugs that inhibit Akt or its downstream effectors may be effective against many human cancers.

To determine how Akt signalling influences tumorigenesis and treatment responses *in vivo*, we compared the effects of a constitutively activated Akt mutant⁹ to the anti-apoptotic regulator Bcl-2 in the E μ -Myc model of B-cell lymphoma¹⁰. E μ -Myc haematopoietic stem cells (HSCs) isolated from fetal livers were transduced with Akt- or Bcl-2-expressing retroviruses and transplanted into lethally irradiated mice¹¹, and the chimaeric recipients were monitored for lymphoma onset and pathology (Fig. 1a). The resulting lymphomas were each transplanted into several wild-type animals, where they produced malignancies that were pathologically indistinguishable from the original disease¹². These animals were then treated with a chemotherapeutic regime and monitored for time to relapse (tumour-free survival) and overall survival by lymph node palpation and blood analysis (Fig. 1b). This approach allows a rapid analysis of genotype–response relationships in a natural setting using methods that parallel human clinical trials.

As with Bcl-2 (ref. 12), Akt markedly accelerated lymphomagenesis relative to controls (Fig. 2a, $P < 0.0001$ for both Akt and Bcl-2 relative to murine stem cell virus (MSCV)). Compared with lymphomas arising in E μ -Myc mice, the Akt and Bcl-2 lymphomas were more invasive and were often associated with leukaemia (Fig. 2d, and data not shown). Immunophenotyping revealed that Akt and Bcl-2 lymphomas were derived from a B220/CD45R-positive progenitor cell that lacked other markers of B-cell differentiation, whereas control lymphomas were typically derived from a more mature B-cell type (Fig. 2b; see also Supplementary Fig. 1 and Supplementary Table 1). Although lymphomas of all genotypes proliferated rapidly as assessed by Ki-67 staining, Akt and Bcl-2 lymphomas showed much less apoptosis relative to controls as measured by a TdT-mediated dUTP nick end labelling (TUNEL) assay (Fig. 2c). The impact of Akt on tumour behaviour was reminiscent of *p53* tumour suppressor gene loss, which accelerates Myc-induced lymphomagenesis by disabling apoptosis¹¹. In fact, whereas all lymphomas arising from *p53*^{+/-} E μ -Myc HSCs lost the wild-type *p53* allele, those expressing either Bcl-2 or Akt did not (Supplementary Fig. 2). Therefore, Akt can compensate for *p53* loss during tumorigenesis and acts in a manner that is temporally and pathologically similar to a strictly anti-apoptotic gene.

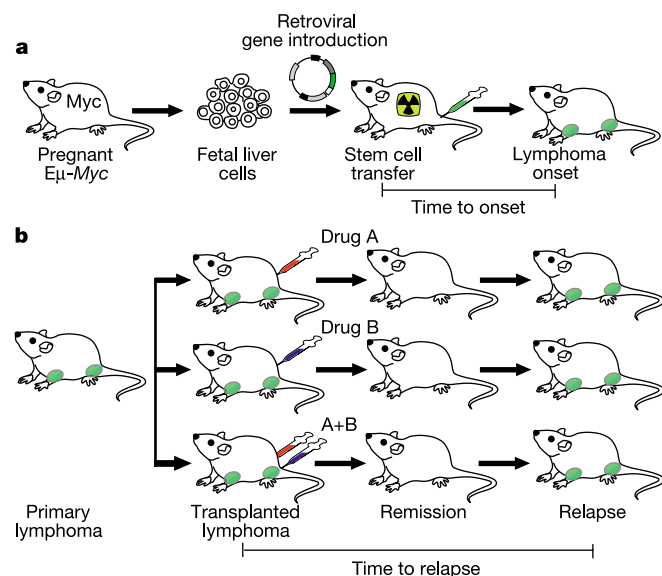


Figure 1 Integrated model of lymphomagenesis and treatment response.

Next, C57BL/6 mice harbouring different transplanted lymphomas were treated with cyclophosphamide (cytoxan, CTX) or doxorubicin (DXR) and monitored for tumour-free and overall survival (Fig. 2e, f; see also Supplementary Fig. 3). Here, animals harbouring *Arf*-null lymphomas were used as controls, as these lymphomas are also highly aggressive but remain chemosensitive¹³. After CTX therapy, mice harbouring control lymphomas invariably entered a complete remission and $>50\%$ remained tumour free after 100 days. In contrast, mice harbouring Akt lymphomas responded poorly, typically displaying remissions of less than 20 days with no long-term survivors ($P < 0.001$ relative to controls). Although DXR produced fewer long-lasting remissions in the control cohort than CTX, mice harbouring Akt lymphomas still displayed substantial reductions in tumour-free and overall survival (Fig. 2f; see also Supplementary Fig. 3b, $P < 0.0001$ for Akt and Bcl-2 versus control). Therefore, Akt-mediated survival signalling can promote drug resistance *in vivo*.

We next investigated whether pharmacologically inhibiting the pathway downstream of Akt might have anti-tumour effects. One established Akt effector is mTOR, a serine/threonine kinase implicated in translation control, which can be potentially inhibited by the immunosuppressant rapamycin¹⁴. Although not considered to be a primary component of Akt survival signalling, mTOR can mediate metabolic changes important for cell survival in growth-factor-poor environments^{15,16}. Rapamycin has had modest anti-tumour activity against PTEN-deficient tumours in animal studies, and can potentiate drug-induced cell death *in vitro*^{17–20}. We therefore asked whether rapamycin was effective against our defined murine lymphomas, and whether its activity was genotype-dependent or enhanced by chemotherapy.

To determine whether rapamycin inhibits mTOR in E μ -Myc lymphomas, its activity was indirectly assessed in extracts from untreated or treated lymphomas using antibodies that specifically recognize the phosphorylated forms of two mTOR targets: the ribosomal S6 protein (which is phosphorylated by the S6 kinase) and eIF4G. As expected, untreated Akt lymphomas expressed substantially more phosphorylated S6 and eIF4G relative to Bcl-2 lymphomas (Fig. 3a, compare lanes 1 and 5). Treatment with rapamycin alone or in combination with chemotherapy substantially reduced S6 and eIF4G phosphorylation in Akt lymphomas (Fig. 3a, compare lanes 1–3 and 4). Hence, rapamycin inhibits mTOR activity *in vivo* in a manner that is not affected by conventional therapies.

We then examined the acute and long-term effects of rapamycin *in vivo*. In Akt lymphomas, rapamycin treatment produced only a slight increase in apoptosis relative to untreated counterparts 6 h after therapy, as assessed by PARP cleavage in lymphoma extracts (Fig. 3a, compare lanes 1 to 3) and by TUNEL staining of lymphoma sections (Fig. 3b). Accordingly, these mice rarely achieved complete remissions (Fig. 4a, b) and they showed tumour-free and overall survival patterns that were no better than those produced by conventional therapy (Fig. 4c; see also Supplementary Fig. 4). Furthermore, rapamycin displayed even less activity against lymphomas of other genotypes (Figs 3 and 4). Therefore, despite its ability to inhibit mTOR *in vivo*, rapamycin had only modest activity as a single agent.

In marked contrast, the combination of chemotherapy and rapamycin showed potent activity against Akt-expressing lymphomas. For example, DXR and rapamycin treatment produced PARP cleavage and massive apoptosis 6 h after therapy (Fig. 3a, compare lanes 1 and 4; see also panel b), with all of the animals achieving complete remissions (Fig. 4a, b; see also Supplementary Fig. 6 for a comparison of matched groups). Indeed, the combination of DXR and rapamycin produced more complete remissions and better tumour-free survival in the otherwise drug-resistant Akt lymphomas than DXR produced in the chemosensitive controls (Fig. 4a, c, $P < 0.001$ for rapamycin plus DXR versus DXR or rapamycin

alone), resulting in a substantial survival benefit (Supplementary Fig. 4a, $P < 0.001$ for rapamycin plus DXR versus DXR or rapamycin alone). Similar effects were observed with the combination of CTX and rapamycin, where approximately half of the mice achieved remissions lasting more than 60 days compared with none for either agent alone (Supplementary Figs 4b and 5a, $P < 0.001$). Consequently, rapamycin reverses the apoptotic defects and drug resistance occurring in Akt lymphomas.

These potent effects were not observed in lymphomas of other genotypes. Bcl-2 lymphomas, which showed a similar drug-resistance profile to either rapamycin or chemotherapy alone, remained non-responsive to a combination of the two treatments as assessed in both acute and long-term assays (Fig. 4a, d; see also Supplementary Figs 4c, d and 5b). Moreover, control lymphomas, which were initially chemosensitive, showed no improvement in short- or long-term responses after combination therapy, and, in fact, rapamycin appeared to antagonize drug-induced apoptosis in these cells (Fig. 3b and our own unpublished observations). Therefore, although each lymphoma studied contained an anti-apoptotic lesion, the chemosensitizing effects of rapamycin were specific for tumours with constitutive Akt signalling.

The results described above suggest that mTOR is essential for Akt-mediated survival signalling. mTOR normally regulates

translation in response to nutrients and growth factors by phosphorylating key components of the protein synthesis machinery, including the above mentioned ribosomal protein S6 kinase, p70^{S6K} and the 4E-BP proteins²¹. Phosphorylation of 4E-BP, in turn, releases the translation initiation factor eIF4E to stimulate cap-dependent translation. Although its role in oncogenesis is poorly understood, eIF4E can have transforming and anti-apoptotic activities *in vitro*^{22,23} and is overexpressed in some tumour types²⁴. Furthermore, Akt-expressing tumours selectively increase the translation of some mRNAs in a rapamycin-reversible manner^{25,26}. Therefore, to examine further the effects of translational control on tumour phenotypes, we directly compared eIF4E to Akt in the μ -Myc model (see Fig. 1).

eIF4E accelerated lymphomagenesis in a manner that was similar to Akt (Fig. 5a, median onset of 42.5 and 50 days for Akt and eIF4E, respectively; $P < 0.0001$ versus control). Moreover, as with Akt lymphomas, eIF4E lymphomas displayed a disseminated pathology and a high proliferation/apoptosis ratio (Fig. 5b), despite only an approximately 2.5-fold increase in eIF4E protein (Fig. 5c, compare lanes 1 to 3; see also Supplementary Fig. 7). eIF4E could also compensate for p53 loss during lymphomagenesis, as lymphomas arising from p53^{+/-} μ -Myc HSCs retained the wild-type p53 allele (Supplementary Fig. 2). Despite these similarities, eIF4E

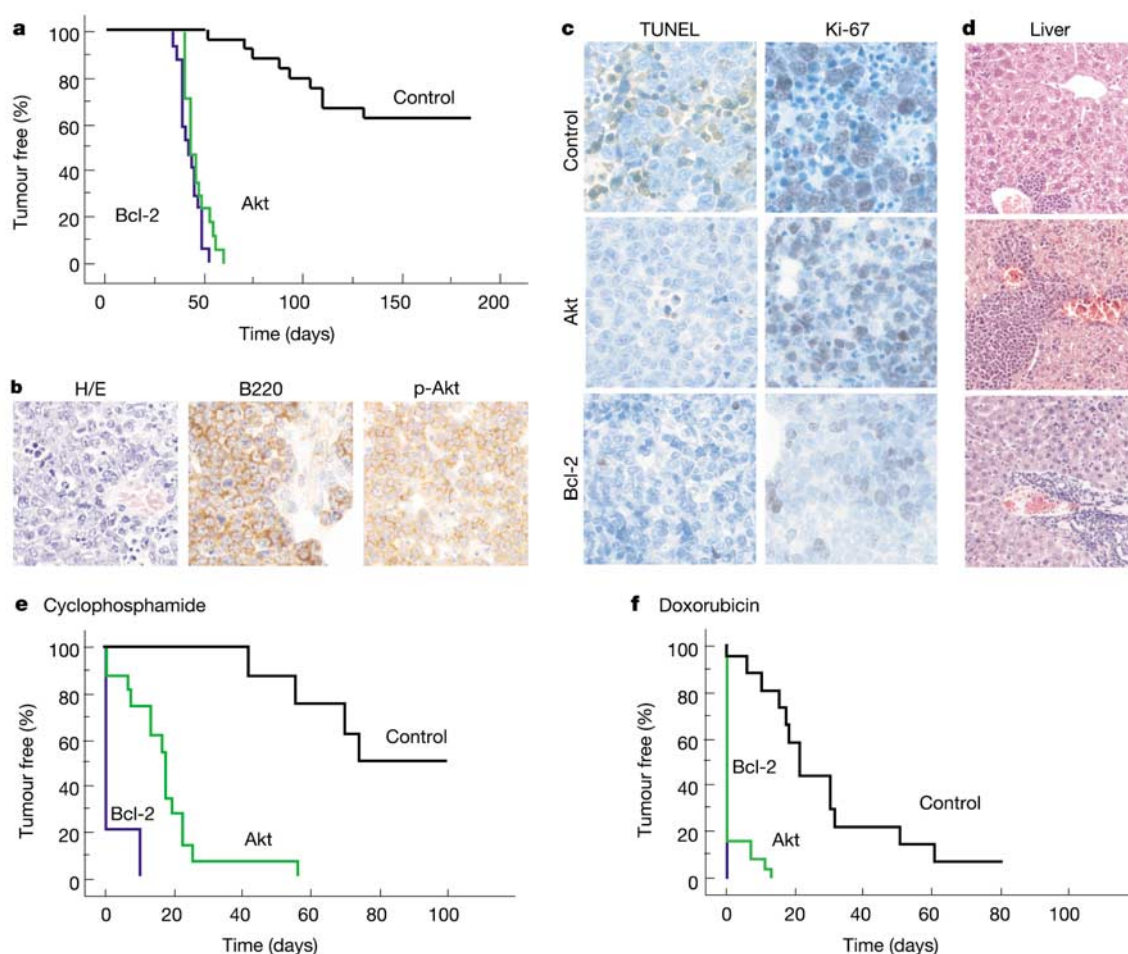


Figure 2 Akt accelerates lymphomagenesis and promotes drug resistance *in vivo*. **a**, Kaplan-Meier plot showing tumour onset in mice reconstituted with μ -Myc HSCs transduced with MSCV ($n = 40$, black), Akt ($n = 18$, green) or Bcl-2 ($n = 18$, blue). A kinase-dead mutant Akt (K179A) did not accelerate lymphomagenesis, and Akt did not cause tumour formation in non-transgenic HSCs (data not shown). **b**, Representative micrographs of μ -Myc/Akt lymphoma sections stained with haematoxylin and eosin (H/E) or the indicated antibody. **c**, TUNEL (left) and Ki-67 (right) staining (brown) of the

indicated lymphoma sections was used to assess apoptosis and proliferation, respectively. **d**, Photomicrographs of haematoxylin-and-eosin-stained liver sections showing perivascular infiltration in control tumours and parenchymal invasion in Akt and Bcl-2 tumours. **e**, **f**, Kaplan-Meier plots showing the tumour-free survival after treatment with CTX (**e**) or DXR (**f**). Mice bearing either control lymphomas (black: DXR, $n = 19$; CTX, $n = 13$), Akt lymphomas (green: DXR, $n = 25$; CTX, $n = 16$) or Bcl-2 lymphomas (blue: DXR, $n = 6$; CTX, $n = 5$) were treated and monitored for time to relapse.

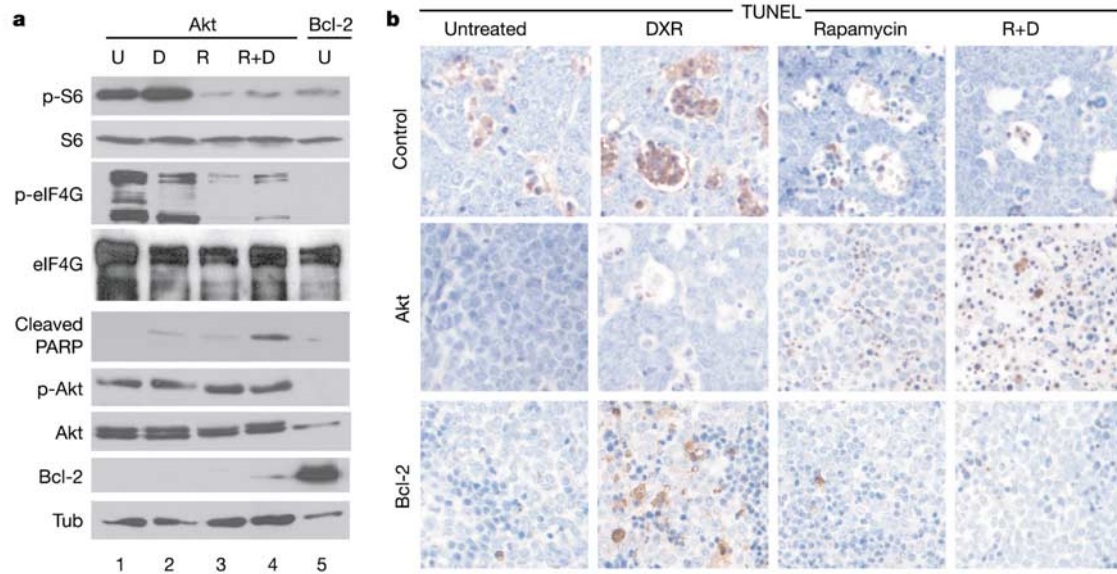


Figure 3 Inhibition of mTOR sensitizes Akt tumours to cytotoxic chemotherapy. Animals harbouring lymphomas of the indicated genotypes were either left untreated (U) or were treated with DXR (D), rapamycin (R) or DXR and rapamycin (R+D), and lymphoma cells were harvested 7 h later. **a**, Lysates were subjected to immunoblotting for phosphorylated

and total ribosomal S6 protein (p-S6 and S6), phosphorylated and total eIF4G (p-eIF4G and eIF4G), Akt (p-Akt and Akt), Bcl-2 and α -tubulin (Tub) as a loading control. **b**, Lymphoma sections derived from untreated or treated animals were stained with TUNEL (brown) to identify apoptotic cells.

lymphomas were derived from a more mature B-cell type and often produced an aggressive infiltration of the renal pelvis distinct from Akt-associated pathologies (Fig. 5b; see also Supplementary Fig. 1). Nevertheless, eIF4E is clearly a potent oncogene *in vivo*, producing phenotypes consistent with an anti-apoptotic gene.

Lymphomas expressing eIF4E were highly resistant to DXR therapy relative to controls, and mice harbouring these lymphomas displayed tumour-free and overall survival patterns that were indistinguishable from those harbouring Akt lymphomas (Fig. 5e,

$P < 0.0001$ versus control; $P = 0.47$ versus Akt). However, in contrast to Akt lymphomas, eIF4E lymphomas displayed no increase in the rapamycin-sensitive phosphorylation of 4E-BP1 or the S6 ribosomal protein (Fig. 5c), and were non-responsive to rapamycin and DXR therapy *in vivo* (Fig. 5e; see also Fig. 4c, $P < 0.0001$). Furthermore, introduction of eIF4E into an initially sensitive Akt lymphoma conferred resistance to rapamycin and chemotherapy; hence, cells co-expressing eIF4E were enriched in mixed populations after therapy relative to those expressing Akt alone (Fig. 5f, g). Together, these results indicate that eIF4E can recapitulate Akt action in oncogenesis and drug resistance. They also show that eIF4E can confer resistance to a rapamycin-based therapy *in vivo*, presumably because it acts downstream of mTOR.

Our results provide new insights into Akt signalling and tumour behaviour. We find that Akt lymphomas are phenotypically similar to those harbouring a purely anti-apoptotic oncogene with respect to their accelerated onset, aggressive pathology, ability to retain p53 and chemotherapy resistance. Thus, despite its pleiotropic activities, the pro-survival functions of Akt seem sufficient to promote tumorigenesis. Furthermore, we show that the mTOR inhibitor rapamycin can restore apoptotic sensitivity to lymphomas expressing Akt, and that the translation factor eIF4E can recapitulate Akt action *in vivo*. Consequently, a substantial proportion of Akt survival signalling may result from deregulated translation, perhaps through altering the recruitment of pro- and anti-apoptotic mRNAs to polysomes^{25,26}. Of note, the apparent role of translational control in cell survival extends to tumours with other PI(3)K pathway lesions, as the chemotherapy resistance of PTEN-deficient lymphomas is reversed by rapamycin in a manner that is also blocked by eIF4E (our own unpublished observations). Together, these results provide *in vivo* evidence that translational signalling through Akt is a broadly relevant oncogenesis and drug-resistance mechanism.

Our results also have important implications for considering the use of targeted therapeutics alone or in combination with conventional agents. By showing that rapamycin can reverse drug resistance in Akt-expressing tumours, we demonstrate that reversing apoptotic defects can restore drug sensitivity *in vivo*. Furthermore, the ability of eIF4E to mediate drug resistance downstream of Akt implies that inhibitors of translation initiation may also be effective

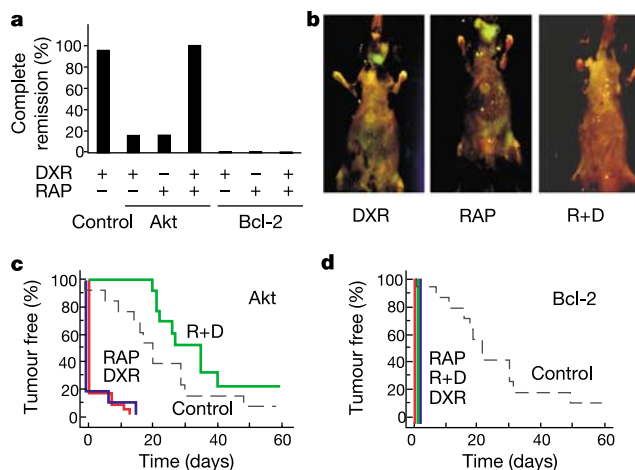


Figure 4 Rapamycin reverses Akt-mediated chemoresistance *in vivo*. **a**, The percentage of mice bearing control ($n = 19$), Akt ($n = 51$) or Bcl-2 ($n = 18$) lymphomas that achieved a complete remission after the indicated treatment. The number of mice in each treatment group ranged from 6 to 25. RAP, rapamycin. **b**, Whole-body fluorescence imaging of GFP expression of a 'matched group' of mice carrying identical Akt tumours 21 days after the indicated treatment. **c, d**, Kaplan-Meier plots of tumour-free survival in Akt lymphomas after treatment with DXR (red, $n = 25$), rapamycin (blue line, $n = 12$) and rapamycin plus DXR (green, R+D, $n = 14$) (**c**), and in Bcl-2 lymphomas after treatment with DXR ($n = 6$, red), rapamycin ($n = 6$, blue) and rapamycin plus DXR (R+D, $n = 6$, green) (**d**). The tumour-free survival of the control tumours after DXR treatment is shown for comparison (dotted line).

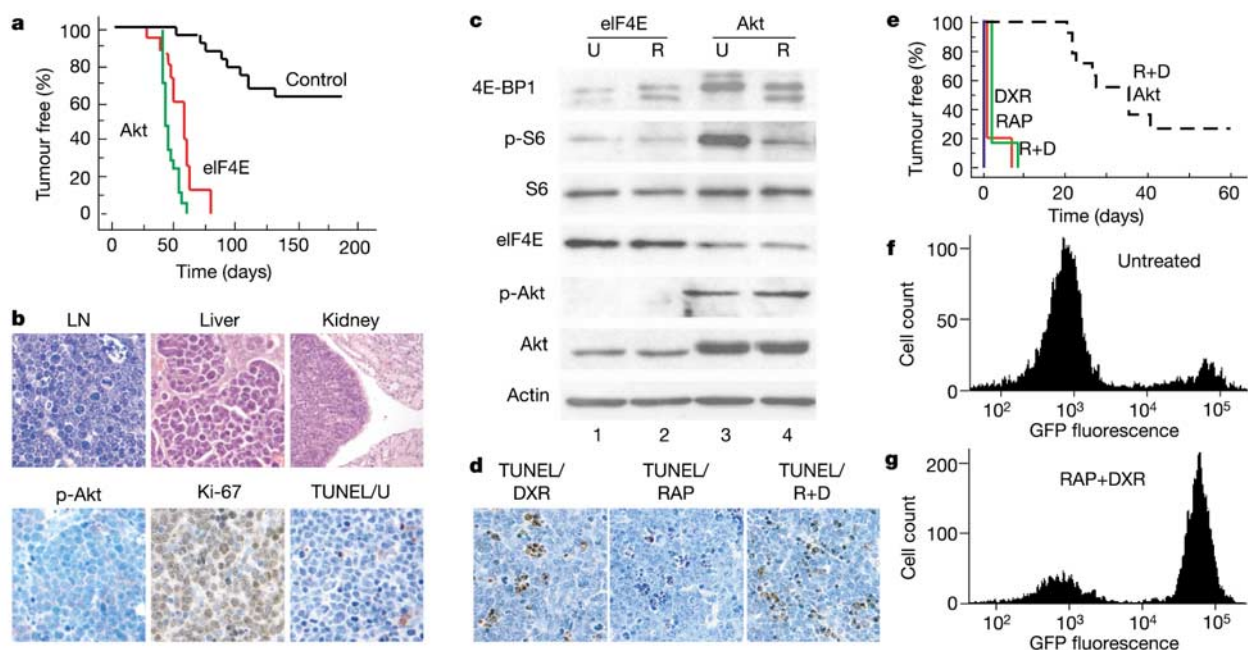


Figure 5 eIF4E promotes oncogenesis and drug resistance *in vivo*. **a**, Kaplan–Meier plots showing tumour onset in mice reconstituted with E μ -Myc HSCs expressing eIF4E (red, $n = 15$). Control and Akt data from Fig. 2 are shown for comparison. eIF4E was unable to induce tumours in normal HSCs within the observation period (> 125 days; data not shown). **b**, Representative micrographs of eIF4E lymphoma sections. Top panel, haematoxylin and eosin staining of the indicated organs; lower panel, immunohistochemical stains for phosphorylated Akt, Ki-67 and TUNEL in untreated tumours. LN, lymph node. **c**, Immunoblotting of lysates derived from eIF4E and Akt lymphomas harvested from untreated (U) animals or 6 h after rapamycin (R) administration, using antibodies against the indicated protein ('p' denotes an antibody specific for a phosphorylated protein). Untreated Akt tumours (lane 3) show an increase in the phosphorylated, slower-migrating forms of

4E-BP1. **d**, TUNEL staining (brown) to assess apoptosis in tumour sections harvested 6–8 h after the indicated therapy. **e**, Kaplan–Meier plots showing tumour-free survival of mice bearing eIF4E lymphomas after treatment with DXR ($n = 5$, red line), rapamycin ($n = 6$, blue) and rapamycin plus DXR ($n = 6$, green line). Data from mice bearing Akt lymphomas that had been treated with rapamycin plus DXR are shown for comparison (dotted line). **f, g**, *In vivo* competition assay. An Akt lymphoma was transduced with MSCV-eIF4E-GFP such that the resulting population contained only about 2–5% GFP-positive cells, and then it was transplanted into multiple recipient mice. Upon lymphoma manifestation, the animals were left untreated (**f**) or treated (**g**) with rapamycin plus DXR therapy. Lymphoma cells were isolated 48 h later and subjected to flow cytometry to determine the fraction of eIF4E-positive cells (high GFP fluorescence).

chemosensitizing agents, and perhaps less prone to resistance. Finally, the selectivity of rapamycin to reverse drug resistance in Akt-expressing lymphomas—despite their broad similarities to those with other apoptotic lesions—implies that the utility of this combination treatment would be missed in clinical trials based solely on tumour type or pathology. Together, our results provide *in vivo* validation for a strategy to reverse drug resistance in human cancers, and underscore the importance of tailoring cancer therapy on the basis of tumour genotype. □

Methods

Generation of lymphomas

E μ -Myc HSCs derived from fetal livers at embryonic day 13–15 were transduced with retroviruses expressing various genes, and used to reconstitute the haematopoietic compartment of lethally irradiated C57BL/6 mice. E μ -Myc $p53^{+/-}$ HSCs were derived from crosses of E μ -Myc and $p53^{+/-}$ mice, and detection of loss of heterozygosity in the $p53$ locus was by allele-specific PCR^{11,27}. The retroviral vectors that were used in this study were MSCV-GFP (the empty vector control), MSCV-Bcl-2 (ref. 12), MSCV-Akt (myrAkt), MSCV-Akt(K179A)⁹ and MSCV-eIF4E. After the appearance of well-palpable lymphomas, tumours were harvested and either fixed for histological evaluation, rendered as single-cell suspensions and frozen in 10% DMSO, or transplanted directly into normal mice for treatment studies¹². For *in vivo* competition experiments, MSCV-eIF4E was transduced into a small percentage of Akt lymphoma cells during a brief *in vitro* passage, and the mixed population was re-injected into several recipient mice. Once lymphomas formed, mice were left untreated or treated with rapamycin and/or DXR. Tumour cells were harvested 48 h later and evaluated for GFP expression by flow cytometry¹².

Treatment studies

A total of 1×10^6 primary lymphoma cells were injected into the tail vein of 6–8-week-old female C57BL/6 mice. After the formation of well-palpable tumours, the animals were treated with rapamycin (4 mg per kg intraperitoneally for 5 days), DXR (10 mg per kg intraperitoneally), CTX (300 mg per kg intraperitoneally) or various combinations. In combination studies, the cytotoxic agent was given on day 2 of the rapamycin protocol.

Rapamycin (LC laboratories) was initially dissolved in 100% ethanol, stored at -20°C , and further diluted in an aqueous solution of 5.2% Tween 80 and 5.2% PEG 400 (final ethanol concentration, 2%) immediately before use. Doxorubicin and cyclophosphamide (both from Sigma) were dissolved in water. In treatment studies, control lymphomas were Arf-null tumours arising in an E μ -Myc/Arf $^{+/-}$ background²⁷. After treatment, the mice were monitored by twice weekly palpation and blood smears.

A complete remission was defined as the absence of detectable tumour and leukaemia. Tumour-free survival was defined as the time between treatment and reappearance of a well-palpable lymphoma¹². Overall survival was defined as the time between treatment and progression to a terminal stage at which the animals were killed. These data were analysed in the Kaplan–Meier format using the log-rank (Mantel–Cox) test for statistical significance. Whole-body fluorescence imaging of living animals was performed as described^{11,28}.

Histopathology

Samples were fixed in 10% buffered formalin and embedded in paraffin. Thin sections (5 μm) were stained with haematoxylin and eosin according to standard protocols. Detection of phosphorylated Akt (rabbit antibody against p-Akt Ser473, 1:100 (Cell Signalling)) and Ki-67 (rabbit antibody, 1:100 (NovoCastra)) was by standard avidin–biotin immunoperoxidase methods, with diaminobenzidine used as the chromogen and haematoxylin as counter stain. For B220 immunohistochemistry (rat antibody against mouse CD45R/B220 clone RA3-6B2 (BD Biosciences, Pharmingen)), antigen retrieval was required. The apoptotic rate was analysed by TUNEL assay according to published protocols²⁹.

Western blot analysis

Immunoblots were performed from whole-cell lysates³⁰. A total of 50 μg of protein/sample were resolved on SDS–PAGE gels and transferred to Immobilon-P membranes (Millipore). Antibodies against phospho-Akt (9275, 1:1,000, Cell Signalling), Akt (9272, 1:1,000, Cell Signalling), ribosomal S6 protein (2212, 1:1,000, Cell Signalling), phospho-ribosomal S6 (2215, 1:1,000, Cell Signalling), cleaved PARP (9548, 1:1,000, Cell Signalling), Bcl-2 (N19, 1:200, Santa Cruz), α -tubulin (B-5-1-2, 1:5,000, Sigma), eIF4E (9742, 1:2,500, Cell Signalling), 4E-BP1 (9452, 1:1,000, Cell Signalling) and β -actin (1:5,000, Sigma) were used as probes and were detected using enhanced chemiluminescence (ECL, Amersham; Lumilight, Roche). eIF4E expression was quantified by immunoblotting using I¹²⁵-labelled recombinant protein A (PerkinElmer) as a secondary reagent, followed by phosphorimager detection.

Received 13 December 2003; accepted 27 January 2004; doi:10.1038/nature02369.

- Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
- Johnstone, R. W., Ruefli, A. A. & Lowe, S. W. Apoptosis: a link between cancer genetics and chemotherapy. *Cell* **108**, 153–164 (2002).
- Mayo, L. D., Dixon, J. E., Durden, D. L., Tonks, N. K. & Donner, D. B. PTEN protects p53 from Mdm2 and sensitizes cancer cells to chemotherapy. *J. Biol. Chem.* **277**, 5484–5489 (2002).
- Datta, S. R., Brunet, A. & Greenberg, M. E. Cellular survival: a play in three Akts. *Genes Dev.* **13**, 2905–2927 (1999).
- Vivanco, L. & Sawyers, C. L. The phosphatidylinositol 3-kinase AKT pathway in human cancer. *Nature Rev. Cancer* **2**, 489–501 (2002).
- Steck, P. A. *et al.* Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nature Genet.* **15**, 356–362 (1997).
- Sakai, A., Thibblemont, C., Wellmann, A., Jaffe, E. S. & Raffeld, M. PTEN gene alterations in lymphoid neoplasms. *Blood* **92**, 3410–3415 (1998).
- Min, Y. H. *et al.* Constitutive phosphorylation of Akt/PKB protein in acute myeloid leukemia: its significance as a prognostic variable. *Leukemia* **17**, 995–997 (2003).
- Andjelkovic, M. *et al.* Role of translocation in the activation and function of protein kinase B. *J. Biol. Chem.* **272**, 31515–31524 (1997).
- Adams, J. M. *et al.* The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* **318**, 533–538 (1985).
- Schmitt, C. A. *et al.* Dissecting p53 tumor suppressor functions *in vivo*. *Cancer Cell* **1**, 289–298 (2002).
- Schmitt, C. A., Rosenthal, C. T. & Lowe, S. W. Genetic analysis of chemoresistance in primary murine lymphomas. *Nature Med.* **6**, 1029–1035 (2000).
- Schmitt, C. A. *et al.* A senescence program controlled by p53 and p16INK4a contributes to the outcome of cancer therapy. *Cell* **109**, 335–346 (2002).
- Huang, S. & Houghton, P. J. Targeting mTOR signalling for cancer therapy. *Curr. Opin. Pharmacol.* **3**, 371–377 (2003).
- Plas, D. R., Talapatra, S., Edinger, A. L., Rathmell, J. C. & Thompson, C. B. Akt and Bcl-xL promote growth factor-independent survival through distinct effects on mitochondrial physiology. *J. Biol. Chem.* **276**, 12041–12048 (2001).
- Edinger, A. L. & Thompson, C. B. Akt maintains cell size and survival by increasing mTOR-dependent nutrient uptake. *Mol. Biol. Cell* **13**, 2276–2288 (2002).
- Neshat, M. S. *et al.* Enhanced sensitivity of PTEN-deficient tumors to inhibition of FRAP/mTOR. *Proc. Natl Acad. Sci. USA* **98**, 10314–10319 (2001).
- Grunwald, V. *et al.* Inhibitors of mTOR reverse doxorubicin resistance conferred by PTEN status in prostate cancer cells. *Cancer Res.* **62**, 6141–6145 (2002).
- Podsypanina, K. *et al.* An inhibitor of mTOR reduces neoplasia and normalizes p70/S6 kinase activity in Pten^{+/-} mice. *Proc. Natl Acad. Sci. USA* **98**, 10320–10325 (2001).
- Hosoi, H. *et al.* Rapamycin causes poorly reversible inhibition of mTOR and induces p53-independent apoptosis in human rhabdomyosarcoma cells. *Cancer Res.* **59**, 886–894 (1999).
- Schmelzle, T. & Hall, M. N. TOR, a central controller of cell growth. *Cell* **103**, 253–262 (2000).
- Lazaris-Karatzas, A., Montine, K. S. & Sonenberg, N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature* **345**, 544–547 (1990).
- Polunovsky, V. A. *et al.* Translational control of the antiapoptotic function of Ras. *J. Biol. Chem.* **275**, 24776–24780 (2000).
- Hershey, J. W. B. & Miyamoto, S. in *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 637–654 (Cold Spring Harbor, New York, 2000).
- Grolleau, A. *et al.* Global and specific translational control by rapamycin in T cells uncovered by microarrays and proteomics. *J. Biol. Chem.* **277**, 22175–22184 (2002).
- Rajasekhar, V. K. *et al.* Oncogenic Ras and Akt signalling contribute to glioblastoma formation by differential recruitment of existing mRNAs to polyosomes. *Mol. Cell* **12**, 889–901 (2003).
- Schmitt, C. A., McCurrach, M. E., de Stanchina, E., Wallace-Brodeur, R. R. & Lowe, S. W. INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. *Genes Dev.* **13**, 2670–2677 (1999).
- Yang, M. *et al.* Whole-body optical imaging of green fluorescent protein-expressing tumors and metastases. *Proc. Natl Acad. Sci. USA* **97**, 1206–1211 (2000).
- Di Cristofano, A., De Acetis, M., Koff, A., Cordon-Cardo, C. & Pandolfi, P. P. Pten and p27KIP1 cooperate in prostate cancer tumor suppression in the mouse. *Nature Genet.* **27**, 222–224 (2001).
- de Stanchina, E. *et al.* E1A signalling to p53 involves the p19(ARF) tumor suppressor. *Genes Dev.* **12**, 2434–2442 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Myers and N. Sonenberg for reagents; C. Rosenthal, M. S. Jiao, P. Chan, M. L. Maunakea and F. Baehner for technical assistance; and L. Bianco for guidance on animal studies. We also thank C. Thompson and members of the Lowe laboratory for discussions, and M. McCurrach, M. Hemann, E. Cepero and D. Burgess for editorial advice. This work was supported by a gift from the Ann L. and Herbert J. Siegel Philanthropic Fund and the Laurie Strauss Leukaemia Foundation, an AACR/Amgen Fellowship in Translational Research (H.-G.W.), a Tularik Post-doctoral Fellowship (E.d.S.), a NSERC graduate scholarship (A.M.), an NCI postdoctoral training grant (J.S.F.), grants from Canadian Institutes of Health Research and National Cancer Institute of Canada (J.P.), the Mouse Models of Human Cancer Consortium and a Burroughs Wellcome Fund Career Award (S.K.), a SCOR grant from the Leukaemia and Lymphoma Society (S.K. and S.W.L.), and a program project grant from the National Cancer Institute (S.W.L. and C.C.-C.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.W.L. (lowe@csh.ledu).

Bmi1 is essential for cerebellar development and is overexpressed in human medulloblastomas

Carly Leung^{1*}, Merel Lingbeek^{2*}, Olga Shakhova¹, James Liu¹, Ellen Tanger², Parvin Saremaslani¹, Maarten van Lohuizen² & Silvia Marino¹

¹Institute of Clinical Pathology, Department of Pathology, University of Zürich, Schmelzbergstrasse 12, 8091 Zürich, Switzerland

²Division of Molecular Genetics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

*These authors contributed equally to this work

Overexpression of the polycomb group gene *Bmi1* promotes cell proliferation and induces leukaemia through repression of *Cdkn2a* (also known as *ink4a/Arf*) tumour suppressors^{1,2}. Conversely, loss of *Bmi1* leads to haematological defects and severe progressive neurological abnormalities in which de-repression of the *ink4a/Arf* locus is critically implicated^{1,3}. Here, we show that *Bmi1* is strongly expressed in proliferating cerebellar precursor cells in mice and humans. Using *Bmi1*-null mice we demonstrate a crucial role for *Bmi1* in clonal expansion of granule cell precursors both *in vivo* and *in vitro*. Deregulated proliferation of these progenitor cells, by activation of the sonic hedgehog (Shh) pathway, leads to medulloblastoma development⁴. We also demonstrate linked overexpression of BMI1 and patched (PTCH), suggestive of SHH pathway activation, in a substantial fraction of primary human medulloblastomas. Together with the rapid induction of *Bmi1* expression on addition of Shh or on overexpression of the Shh target *Gli1* in cerebellar granule cell cultures, these findings implicate BMI1 overexpression as an alternative or additive mechanism in the pathogenesis of medulloblastomas, and highlight a role for *Bmi1*-containing polycomb complexes in proliferation of cerebellar precursor cells.

We characterized the expression of *Bmi1* during mouse cerebellar development (at embryonic day (E)14.5 and E16.5 and at postnatal day 1, 5, 8, 15 and 30). Immunohistochemical analysis showed strong nuclear staining for *Bmi1* in the cells of the external granular layer (EGL) (Fig. 1, upper row; see also Supplementary Fig. S1) and weaker staining in neural precursor cells of the cerebellar neuro-epithelium (Fig. 1, upper row) at all embryonic time points analysed. During postnatal development the strongest *Bmi1* staining was observed between days 5 and 8 in actively proliferating granule cell precursors located in the EGL (Fig. 1, middle panel of left column). Immunoblotting of extracts obtained from EGL precursor cells and/or granule cells of newborn mice and mice at postnatal day 5, 8 and 15 confirmed high *Bmi1* expression in proliferating granule cell precursors and showed its downregulation in postmitotic and terminally differentiated granule cells (Fig. 1, middle panel of right column). Notably, the time course of this expression paralleled the expression of N-Myc and its target gene cyclin D2 (Fig. 1, middle panel of right column), which are involved in the rapid expansion of progenitor cells during neurogenesis^{5–7}. Immunohistochemical staining on autaptic cerebellar sections of human fetuses at gestational week 17, 19, 21, 26 and 32, and of a 2-month-old child and an adult, demonstrate a comparable expression pattern (Fig. 1, bottom panels).

In addition to haematopoietic and skeletal defects, *Bmi1*-deficient mice develop progressive ataxic gait, balance disorders, tremors and behavioural abnormalities at the age of 2–4 weeks³. Although the mice survive to adulthood, they exhibit severe reduction in total postnatal brain mass with particularly severe

phenotypic changes in the cerebellum, as expected from the clinical signs (ref. 3; see also Fig. 2a).

Histological analysis revealed normal cerebellar architecture but a markedly reduced cellularity in the granular and molecular layers (Fig. 2a, NeuN and parvalbumin, and c, area/cell density measurements), in addition to an almost complete lack of stellate neurons and striking morphological abnormalities of the dendritic arborizations of the few basket neurons present (Fig. 2a, NF200). Immunohistochemical staining for NeuN (granule cells), calbindin and parvalbumin (Purkinje cells), as well as messenger RNA *in situ* hybridization for metabotropic glutamate receptor 2 (mGluR-2; Golgi cells, data not shown), revealed normal morphology and terminal differentiation of these cerebellar neurons (Fig. 2a).

To track down the defects in cerebellar development in *Bmi1*-deficient mice, we analysed E14.5 and E16.5 mouse embryos as well as mice at postnatal day 1, 8 and 15. The overall size of the cerebellum is reduced from E16.5 onwards, and the *Bmi1*-null EGL and neuroepithelium appear 1–2 cells thinner (Fig. 2b). In line with a critical role of *Bmi1* in proliferation of EGL progenitor cells, we found a significant reduction of MATH-1-expressing mitotically active granule cell progenitors in the outer EGL of newborn *Bmi1*^{-/-} mice. Widespread expression of p27 throughout the whole width of the EGL confirmed the postmitotic state of most of the *Bmi1*^{-/-} EGL cells (Fig. 2b). At the peak of EGL proliferation at postnatal day 8, *Bmi1*^{-/-} mice showed a significant reduction of 5-bromodeoxyuridine (BrdU)-positive EGL cells (Fig. 2c) as well as

a marked increase in apoptotic cells (Fig. 2c), which might imply that failure to respond to proliferation induction shunts EGL cells into cell death.

Together these results demonstrate a crucial role of *Bmi1* in the maintenance and expansion of immature granule cell precursors. This extends the recently reported essential role of *Bmi1* in renewal of adult haematopoietic stem cells^{8,9} and cortical neural stem cells (ref. 10 and D. Zencak *et al.*, unpublished work) to include more

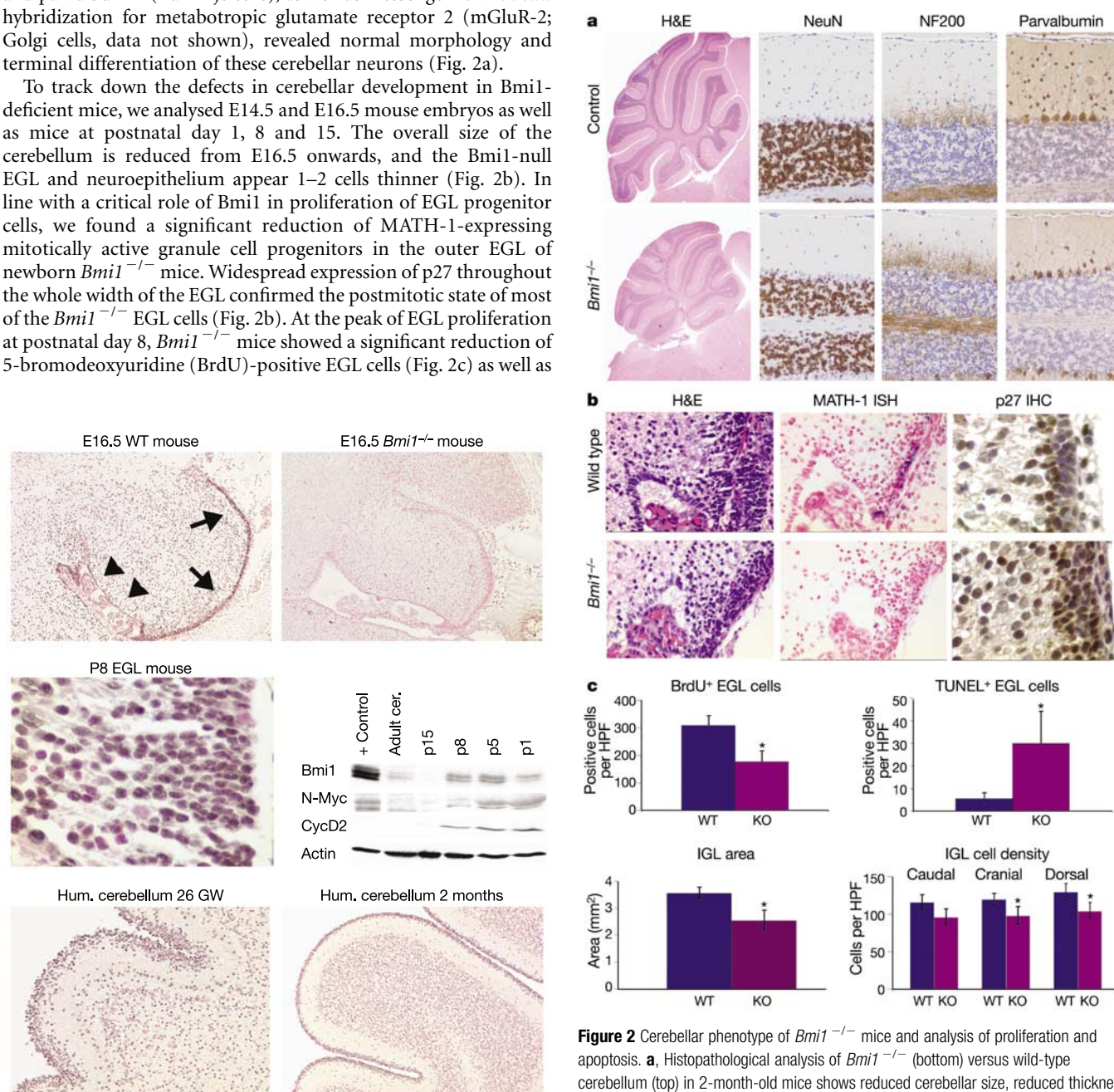


Figure 2 Cerebellar phenotype of *Bmi1*^{-/-} mice and analysis of proliferation and apoptosis. **a**, Histopathological analysis of *Bmi1*^{-/-} (bottom) versus wild-type cerebellum (top) in 2-month-old mice shows reduced cerebellar size, reduced thickness of molecular and granular layers and lack of stellate cells in the *Bmi1*^{-/-} mouse. H&E, haematoxylin and eosin. Final magnification, $\times 5$, $\times 60$. **b**, Reduction of MATH-1-positive EGL precursors and wider expression of the postmitotic marker p27 in *Bmi1*^{-/-} mice at postnatal day 1. IHC, immunohistochemistry; ISH, *in situ* hybridization. Final magnification, $\times 80$, $\times 160$. **c**, Significant reduction of BrdU incorporation (asterisk, $P < 0.001$) and increased apoptosis (asterisk, $P < 0.006$) are observed in the EGL of *Bmi1*^{-/-} mice at postnatal day 8. Area measurements and cell density counts show significant (asterisk, $P < 0.008$ and $P < 0.05$) reduction of size and cell density of the internal granular layer (IGL) as early as postnatal day 15. HPF, high-power field; KO, knockout (*Bmi1*^{-/-}); WT, wild type.

Figure 1 *Bmi1* expression in cerebellar precursor cells during development. *Bmi1* is expressed in EGL precursors (arrows) during embryonic (E16.5 and gestational week (GW) 26) and early postnatal (P) (8 days and 2 months) cerebellar development, most notably during clonal expansion, in mouse (top and middle rows) and human (bottom row). Weaker expression is found in neuroepithelium precursors (arrowheads) and in terminally differentiated neurons (immunohistochemical analysis, *Bmi1* F6 antibody). Final magnification, $\times 28$, $\times 56$ and $\times 38$. Western blot analysis of granule cell precursors/granule cells isolated at postnatal day 1, 5, 8 and 15 (middle panel of right column) confirmed the expression of *Bmi1* in proliferating granule cell precursors.

committed cerebellar precursor cells. No gross impairment of fate determination and terminal differentiation of neural precursor cells was observed in the absence of *Bmi1*.

Cerebellar granule cell (CGC) cultures, which can be induced to proliferate on addition of Shh⁴¹¹, are a well-established method to study neural precursor cell proliferation and differentiation *in vitro*. We show here that *Bmi1* mRNA is upregulated in response to Shh treatment as early as 4 h after incubation (Fig. 3a). Expression of *Bmi1* consistently remains high for as long as 72 h after incubation with Shh (Fig. 3a, b), and parallels *Gli1*, suppressor of fused (*Sufu*) and cyclin D2 expression, indicative of activation of the Shh pathway and induction of proliferation. Moreover, overexpression of *Gli1* in CGCs and in rat kidney cells induces *Bmi1* expression (Fig. 3c), implying that *Bmi1* is a downstream target in the Shh pathway. As confirmation of this, we found a significant reduction in proliferation, as measured by BrdU incorporation, after Shh treatment in CGCs lacking *Bmi1* when compared with wild-type CGCs (Fig. 3d).

Next, we overexpressed *Bmi1* in *Bmi1*^{-/-} CGCs in the absence of Shh. As is shown in Fig. 3e, f, *Bmi1* overexpression induces proliferation, as measured by induction of the DNA replication marker PCNA and by increased BrdU incorporation in *Bmi1*^{-/-} CGCs. In contrast to massive induction of cyclin D2 expression by Shh, overexpression of *Bmi1* in *Bmi1*^{-/-} CGCs does not

significantly increase cyclin D2 levels, and the effects on BrdU incorporation are reduced when compared with Shh stimulation. This finding is most compatible with a dual and separate action of the Shh pathway on both *Bmi1* and N-Myc/cyclin D2 (see below). The similar expression kinetics of *Bmi1* and known downstream targets of the Shh pathway during granule cell development *in vivo*; their critical role in granule precursor cell proliferation; and the similar cerebellar phenotype of mice lacking N-Myc⁶, cyclin D2 (ref. 7) or *Bmi1* (ref. 3), strongly support a critical role of *Bmi1* in Shh-dependent proliferation of granule cell progenitors.

Medulloblastomas—highly malignant brain tumours of childhood—originate from an uncontrolled proliferation of EGL precursor cells and represent an intriguing example of neoplastic transformation of undifferentiated progenitor cells¹². To assess whether BMI1 might be involved in the pathogenesis of human medulloblastomas, we analysed three medulloblastoma cell lines (DAOY, D-341 and D-458) and 12 primary human medulloblastoma samples for BMI1 expression. Western blot analysis revealed strong overexpression of BMI1 in all cell lines (Fig. 4b) and in eight out of 12 primary tumours tested (Fig. 4c) when compared with normal cerebellum. CDK2 protein levels, a marker for proliferation, served as a control to exclude significant contamination of normal tissue, and was found to be equally expressed in all

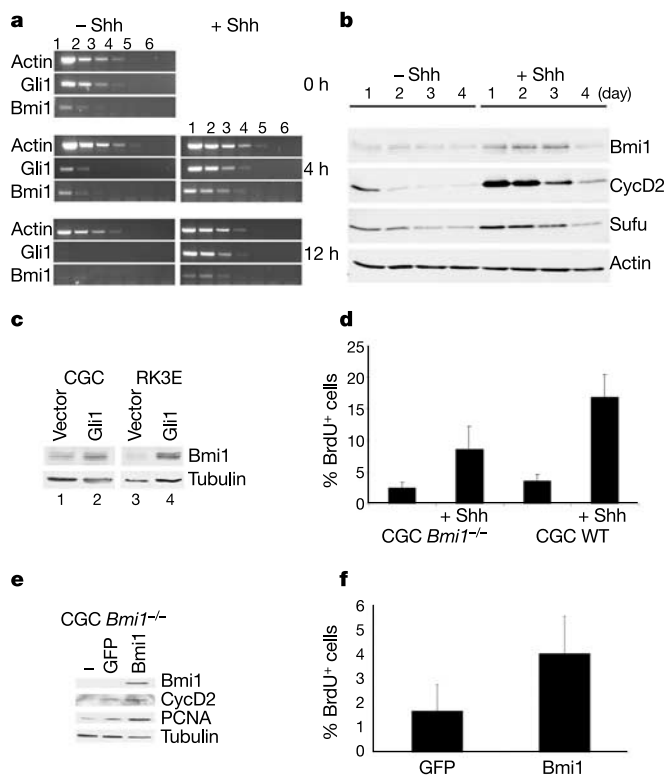


Figure 3 *Bmi1* is essential for efficient CGC proliferation on Shh treatment. **a**, Upregulation of *Bmi1* expression parallels upregulation of *Gli1* on Shh treatment. Semi-quantitative RT-PCR analysis at 0, 4 and 12 h with and without Shh treatment. Lanes 1–5, serial cDNA dilutions; lane 6, control. **b**, Time course of *Bmi1*, cyclin D2 and *Sufu* expression at later time points after Shh treatment (western blot). **c**, *Gli1* overexpression induces upregulation of *Bmi1* in wild-type CGCs and in RK3E cells. **d**, Reduced BrdU incorporation in *Bmi1*^{-/-} CGCs versus wild-type CGCs on Shh treatment ($P < 0.003$). **e**, Lentiviral overexpression of *Bmi1* in *Bmi1*^{-/-} CGCs induces the proliferation marker PCNA. No induction of cyclin D2 is observed. **f**, Overexpression of *Bmi1* in *Bmi1*^{-/-} CGCs increases BrdU incorporation ($P < 0.02$). The weak induction of proliferation observed in the control infection with GFP (**e**, **f**) is probably due to serum components co-precipitating during the ultracentrifugation step during lentivirus preparation.

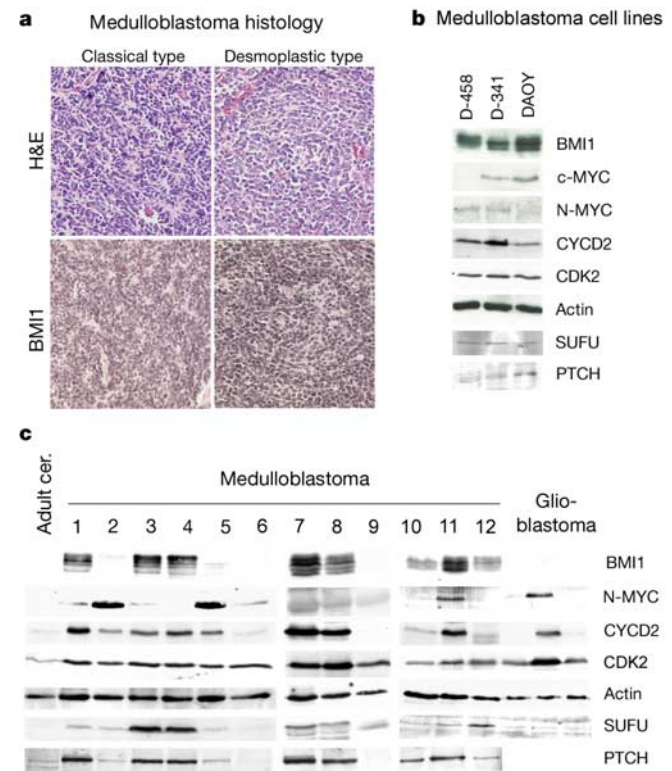


Figure 4 Overexpression of BMI1 in 67% of analysed medulloblastomas correlates with SHH pathway activation. **a**, Both classical and desmoplastic medulloblastomas overexpress BMI1. Final magnification, $\times 50$. **b**, Overexpression of BMI1 correlates with overexpression of cyclin D2 and c-MYC in medulloblastoma cell lines. No overexpression of N-MYC is observed. Equal expression of CDK2 (proliferation marker) and actin is shown. **c**, Overexpression of BMI1 shows a marked correlation with overexpression of PTCH and cyclin D2 in medulloblastomas (samples 1, 3, 4, 7, 8, 10, 11 and 12). Co-overexpression of N-MYC and BMI1 is observed in two tumour samples. Equal expression of CDK2 and actin is shown. Note that tumours without BMI1 overexpression (2, 5, 6 and 9) have low or absent levels of PTCH. Overexpression of BMI1 is specific to medulloblastomas, as three analysed glioblastomas do not show high BMI1 levels (see also Supplementary Fig. S2).

samples (Fig. 4c). In contrast, ten glioblastomas—highly proliferating brain tumours of glial origin that are not related to medulloblastomas—showed no overexpression of BMI1 (Fig. 4c; see also Supplementary Fig. S2).

Immunohistochemical analysis of primary medulloblastomas revealed no significant correlation between high BMI1 expression and the histological variant of the tumours (classical or desmoplastic; Fig. 4a).

Mutation analysis in human medulloblastomas, as well as analysis of the development of medulloblastomas in a fraction of mice heterozygous for a null mutation of *Ptch*, suggests that inappropriate maintenance of Shh signalling in proliferative EGL precursor cells promotes medulloblastoma formation. Indeed mutations of *PTCH*, smoothened (*SMO*H) and *SUFU*, leading to activation of the SHH pathway, have been found in 25% of sporadic human medulloblastomas^{13–15}.

Overexpression of BMI1 correlated well with overexpression of *PTCH* and *SUFU* (Fig. 4c), suggesting at least a partial activation of the hedgehog pathway in BMI1-overexpressing tumours. Notably, all tumours with low levels of *PTCH* expression did not overexpress BMI1, further implying that BMI1 is a possible downstream effector of SHH signalling and that BMI1 overexpression is an alternative or additive mechanism to *PTCH* mutation in medulloblastoma pathogenesis. The observed frequent overexpression of *PTCH* protein is in agreement with previous analyses of mRNA levels^{16,17} but it contrasts with mRNA expression profiling demonstrating *PTCH* activation mainly in desmoplastic medulloblastomas¹⁸. Although the reason for this remains to be clarified, it may in part be attributable to different detection methods (protein versus mRNA profile analysis).

Bmi1 actively collaborates with the *c-Myc* oncogene in murine lymphomagenesis through inhibition of Myc-induced apoptosis via downregulation of *ink4a/ARF* (ref. 2). Moreover, N-Myc has recently been shown to be essential for the rapid expansion of progenitor cell populations during neurogenesis⁶, and to be regulated by hedgehog signalling. Amplification of *c-MYC*, and more often of *N-MYC*, occurs in human medulloblastomas and correlates with poor clinical outcome^{19,20}. As shown in Fig. 4b, *c-MYC* is overexpressed in two out of three medulloblastoma cell lines but not in the human tumour samples; however, we found overexpression of N-MYC in four out of 12 medulloblastomas, of which two also demonstrated BMI1 overexpression (Fig. 4c). These data suggest that BMI1 is not a target of N-MYC.

It is possible that the SHH pathway branches upstream of N-MYC, and that both Bmi1 induction and N-Myc upregulation (leading to cyclin D1 and D2 induction⁵) by Gli transcription factors are required for full Shh-mediated proliferation response of CGCs. Bmi1 affects the activity of cyclin D, Cdk4/Cdk6 and p53 by repressing *p16^{ink4a}* and *p19^{arf}* tumour suppressors¹. As such, BMI1 overexpression in a majority of human medulloblastomas can potentially lead to repression of both the RB and p53 pathways, thus disturbing the balance between differentiation and proliferation of cerebellar precursors. Although we cannot exclude the involvement of other Bmi1 targets, it is intriguing in this respect that medulloblastomas develop with high frequency in mice selectively lacking Rb and p53 in EGL precursors²¹, whereas loss of these two tumour suppressors are rare events in the pathogenesis of human medulloblastomas²².

The recently established requirement for Bmi1 in controlling proliferation of both normal and leukaemic adult stem cells^{8,9,23}, together with the essential role of Bmi1 in the proliferation of EGL progenitor cells and its overexpression in a majority of medulloblastomas (this paper), indicate a common conserved role for Bmi1-containing polycomb complexes in the maintenance and expansion of stem cells or committed progenitors, and in the pathogenesis of tumours originating from the neoplastic transformation of these cells. □

Methods

Isolation of CGCs and culturing conditions

CGC cultures were isolated from 7–8-day-old FVB mice according to established protocols¹¹. After isolation, 1.6×10^6 cells ml^{-1} were seeded onto plates pre-coated with $100 \mu\text{g ml}^{-1}$ poly-L-lysine (Sigma) and allowed to adhere for 30 min at 37 °C, 5% CO_2 . Non-adherent cells were removed after 30 min and replaced with fresh media.

For proliferation experiments, CGCs were cultured in neural basal medium (Gibco) containing 2 mM L-glutamine, $4 \mu\text{g ml}^{-1}$ insulin, 1 mM sodium pyruvate, $1 \times$ Penicillin/Streptomycin (Gibco), $0.062 \mu\text{g ml}^{-1}$ progesterone, $16 \mu\text{g ml}^{-1}$ putrescine, $100 \mu\text{g ml}^{-1}$ BSA, $0.04 \mu\text{g ml}^{-1}$ sodium selenite, $100 \mu\text{g ml}^{-1}$ apo-transferrin and 25 mM KCl. All chemicals were obtained from Sigma. For Shh induction of proliferation, cells were incubated with $3 \mu\text{g ml}^{-1}$ of Shh (R&D), added immediately after media change and harvested at 4, 12, 24, 48, 72 and 96 h for RNA and protein analysis.

For BrdU incorporation assays, CGCs were pulsed with $25 \mu\text{g ml}^{-1}$ BrdU for 2 h before fixation in 4% paraformaldehyde.

E1A-immortalized rat kidney RK3E cells are susceptible to transformation by Gli1 (ref. 24). Gli1 complementary DNA was cloned into a MCSV retrovirus and used to infect RK3E cells according to standard protocols¹.

Immunoblotting and RT-PCR experiments

Protein analysis was performed according to standard protocols. A list of the antibodies and dilutions used is available as Supplementary Information.

Total RNA was isolated using the RNeasy Mini kit (Qiagen). One microgram of total RNA was reverse transcribed using Superscript III RT and oligo-dT primers (Invitrogen). cDNA was diluted 5, 25, 125, 625 and 3,125 times and $2 \mu\text{l}$ was used as template for further polymerase chain reaction (PCR) amplification. Amplification was performed as follows: 94 °C for 3 min followed by 35 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 1 min. All primer sequences are listed in the Supplementary Information.

CGC infection with lentivirus

Gli1 cDNA was cloned in the HIV-CS-CG transfer vector, replacing the green fluorescent protein (GFP) marker gene. Lentivirus for Bmi or Gli1 overexpression was produced as previously described²⁵ and was added to freshly established CGC cultures 12 h after seeding. Infected CGCs were harvested 72 h after infection.

Histology, immunohistochemistry and *in situ* hybridization

For routine sections, brains were removed, immersion-fixed in 4% buffered paraformaldehyde for at least 4 h, cut in coronal or sagittal planes, paraffin embedded, cut in 3- μm -thick sections, and stained with haematoxylin and eosin. A Ventana Nexes Machine was used for immunohistochemical staining. A list of the antibodies and dilutions used is available as Supplementary Information. TdT-mediated dUTP nick end labelling (TUNEL) staining was performed with the *in situ* cell death kit (Roche).

For BrdU detection in cells, coverslips with fixed CGCs were treated for 2 min with 2 M HCl and incubated with mouse anti-BrdU antibody (DAKO) and with fluorescein-isothiocyanate-conjugated anti-mouse antibody. DAPI (4,6-diamidino-2-phenylindole) staining was used to label nuclei. The percentage of BrdU-positive cells on the total amount of DAPI-stained cells was calculated. Four 400-cell fields were counted per condition, using separate litters of animals and two coverslips per isolate.

In situ hybridizations for proteolipid protein, mGluR-2 and MATH-1 were carried out as described in ref. 21. The TSA amplification kit was used for mGluR-2.

Cell counts

Area measurement of the internal granular layer was carried out with AnalySiS imaging software. Cell density was determined in cranial, dorsal and caudal cerebellar regions (total number of internal granular layer neurons in three high-power fields each) for three wild-type and three *Bmi1*^{-/-} mice (postnatal day 15), respectively. BrdU-positive or TUNEL-positive cells were counted in ten high-power fields centred on the EGL in four wild-type and four *Bmi1*^{-/-} mice (postnatal day 8), respectively.

Received 4 November 2003; accepted 2 February 2004; doi:10.1038/nature02385.

- Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A. & van Lohuizen, M. The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus. *Nature* **397**, 164–168 (1999).
- Jacobs, J. J. *et al.* Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via *INK4a/ARF*. *Genes Dev.* **13**, 2678–2690 (1999).
- van der Lugt, N. M. *et al.* Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes Dev.* **8**, 757–769 (1994).
- Dahmane, N. & Ruiz-i-Altaba, A. Sonic hedgehog regulates the growth and patterning of the cerebellum. *Development* **126**, 3089–3100 (1999).
- Kenney, A. M., Cole, M. D. & Rowitch, D. H. Nmyc upregulation by sonic hedgehog signaling promotes proliferation in developing cerebellar granule neuron precursors. *Development* **130**, 15–28 (2003).
- Knoepfler, P. S., Cheng, P. F. & Eisenman, R. N. N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation. *Genes Dev.* **16**, 2699–2712 (2002).
- Huard, J. M., Forster, C. C., Carter, M. L., Sicinski, P. & Ross, M. E. Cerebellar histogenesis is disturbed in mice lacking cyclin D2. *Development* **126**, 1927–1935 (1999).
- Park, I. K. *et al.* Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302–305 (2003).
- Lessard, J. & Sauvageau, G. *Bmi-1* determines the proliferative capacity of normal and leukaemic stem cells. *Nature* **423**, 255–260 (2003).

10. Molofsky, A. V. *et al.* Bmi1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**, 962–967 (2003).
11. Wechsler-Reya, R. J. & Scott, M. P. Control of neuronal precursor proliferation in the cerebellum by Sonic Hedgehog. *Neuron* **22**, 103–114 (1999).
12. Rubin, J. B. & Rowitch, D. H. Medulloblastoma: a problem of developmental biology. *Cancer Cell* **2**, 7–8 (2002).
13. Zurawel, R. H. *et al.* Analysis of PTCH/SMO/SHH pathway genes in medulloblastoma. *Genes Chromosomes Cancer* **27**, 44–51 (2000).
14. Wechsler-Reya, R. & Scott, M. P. The developmental biology of brain tumors. *Annu. Rev. Neurosci.* **24**, 385–428 (2001).
15. Taylor, M. D. *et al.* Mutations in SUFU predispose to medulloblastoma. *Nature Genet.* **31**, 306–310 (2002).
16. Dong, J., Gailani, M. R., Pomeroy, S. L., Reardon, D. & Bale, A. E. Identification of PATCHED mutations in medulloblastomas by direct sequencing. *Hum. Mutat.* **16**, 89–90 (2000).
17. Pietsch, T. *et al.* Medulloblastomas of the desmoplastic variant carry mutations of the human homologue of *Drosophila* patched. *Cancer Res.* **57**, 2085–2088 (1997).
18. Pomeroy, S. L. *et al.* Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **415**, 436–442 (2002).
19. Herms, J. *et al.* C-MYC expression in medulloblastoma and its prognostic value. *Int. J. Cancer* **89**, 395–402 (2000).
20. Tomlinson, F. H. *et al.* Aggressive medulloblastoma with high-level N-myc amplification. *Mayo Clin. Proc.* **69**, 359–365 (1994).
21. Marino, S., Vooijs, M., van Der Gulden, H., Jonkers, J. & Berns, A. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes Dev.* **14**, 994–1004 (2000).
22. Lee, Y. *et al.* A molecular fingerprint for medulloblastoma. *Cancer Res.* **63**, 5428–5437 (2003).
23. Smith, K. S. *et al.* Bmi-1 regulation of INK4A-ARF is a downstream requirement for transformation of hematopoietic progenitors by E2a-Pbx1. *Mol. Cell* **12**, 393–400 (2003).
24. Ruppert, J. M., Vogelstein, B. & Kinzler, K. W. The zinc finger protein GLI transforms primary cells in cooperation with adenovirus E1A. *Mol. Cell Biol.* **11**, 1724–1728 (1991).
25. Miyoshi, H., Blomer, U., Takahashi, M., Gage, F. H. & Verma, I. M. Development of a self-inactivating lentivirus vector. *J. Virol.* **72**, 8150–8157 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Kleihues and H. Ohgaki for providing medulloblastoma samples; M. Grozer for medulloblastoma cell lines; E. R. Fearon for RK3E cells; M. Ruppert for Gli cDNAs; D. Trono for the lentivirus Bmi1 construct; I. Camenisch for technical help; and K. Kieboom for animal care. MATH-1 and mGluR-2 probes were gifts from H. Zoghbi and S. Nakanishi, respectively. We thank A. Lund, M. Hernandez and S. Bruggeman for discussions, and P. U. Heitz for support. This work was supported by grants from the 'Krebsforschung Schweiz' to S.M. and from the 'Novartis Stiftung' to S.M.; M. L. and E. T. were supported by a Pioneer grant from the Netherlands organization for Scientific Research to M.v.L.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.M. (silvia.marino@usz.ch) or M.v.L. (m.v.lohuizen@nki.nl).

The protein kinase PKR is required for macrophage apoptosis after activation of Toll-like receptor 4

Li-Chung Hsu¹, Jin Mo Park¹, Kezhong Zhang³, Jun-Li Luo¹, Shin Maeda¹, Randal J. Kaufman³, Lars Eckmann², Donald G. Guiney² & Michael Karin¹

¹Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology, and ²Department of Medicine, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0636, USA
³Howard Hughes Medical Institute, Department of Biological Chemistry, University of Michigan, 1150 W. Medical Center Drive, Ann Arbor, Michigan 48109-0659, USA

Macrophages are pivotal constituents of the innate immune system, vital for recognition and elimination of microbial pathogens¹. Macrophages use Toll-like receptors (TLRs) to detect pathogen-associated molecular patterns—including bacterial cell wall components, such as lipopolysaccharide or lipoteichoic acid, and viral nucleic acids, such as double-stranded (ds)RNA—and in turn activate effector functions, including anti-apoptotic

signalling pathways². Certain pathogens, however, such as *Salmonella* spp., *Shigellae* spp. and *Yersiniae* spp., use specialized virulence factors to overcome these protective responses and induce macrophage apoptosis³. We found that the anthrax bacterium, *Bacillus anthracis*, selectively induces apoptosis of activated macrophages⁴ through its lethal toxin, which prevents activation of the anti-apoptotic p38 mitogen-activated protein kinase⁴. We now demonstrate that macrophage apoptosis by three different bacterial pathogens depends on activation of TLR4. Dissection of anti- and pro-apoptotic signalling events triggered by TLR4 identified the dsRNA responsive protein kinase PKR as a critical mediator of pathogen-induced macrophage apoptosis. The pro-apoptotic actions of PKR are mediated both through inhibition of protein synthesis and activation of interferon response factor 3.

At least ten TLRs are known, and some of the pathogen-associated molecular patterns that cause their activation were identified². Lethal toxin (LT) or p38 inhibitors induce apoptosis in macrophages incubated with either lipopolysaccharide (LPS) or lipoteichoic acid (LTA), derived from Gram-negative and Gram-positive bacteria, respectively⁴. We have now found that heat-inactivated *B. anthracis*, a Gram-positive bacterium, induces extensive macrophage apoptosis in the presence of SB202190, a p38 inhibitor, but bone-marrow-derived macrophages (BMDMs) from C3H/HeJ mice, whose TLR4 is inactive⁵, are resistant to such killing (Fig. 1a). The apoptotic response to heat-killed *B. anthracis* was also seen with bacteria grown in LPS-free medium (data not shown). In wild-type BMDMs, a strong apoptotic response dependent on p38 inhibition was detected only upon treatment with the TLR4 agonist, LPS (Fig. 1b). Little or no apoptosis was seen after incubation with the TLR2 agonist synthetic bacterial lipopeptide (Pam₃CSK₄) or the TLR9 agonist immunostimulatory (CpG-containing) DNA. The TLR3 agonist, synthetic dsRNA—poly(IC)—induced a weak apoptotic response even without p38 inhibition. Transient expression of a CD4–hToll chimera⁶, in which the intracellular Toll-IL-1 receptor (TIR) domain of TLR4 was fused to the extracellular and trans-membrane domains of CD4, also resulted in apoptosis after p38 inhibition (Fig. 1c). Consistent with the critical role of TLR4, BMDMs from C3H/HeJ mice, but not from the equivalent wild-type strain, C3H/HeOuJ, were resistant to apoptosis induced by LPS plus SB202190 (Fig. 1d).

TLR4 uses several adaptor proteins, including MyD88, MAL/TIRAP, TRIF and TRAM to engage downstream signalling proteins and eventually activate IκB kinase (IKK) and mitogen-activated protein kinases (MAPKs)^{7,8} (Fig. 2a). Macrophages from mice lacking MyD88 (ref. 9) or TRAF6, a signalling protein that acts downstream of MyD88 and TRAP¹⁰, still undergo apoptosis after LPS stimulation and p38 inhibition (Fig. 2b, c). In fact, both MyD88^{−/−} and TRAF6^{−/−} macrophages exhibit an increased apoptotic response. NF-κB activation and IκB degradation, which depend on IKKβ¹¹, are reduced in both MyD88^{−/−} and TRAF6^{−/−} macrophages (Fig. 2d and data not shown). As NF-κB activates anti-apoptotic genes¹¹, these defects may explain the enhanced apoptosis of MyD88^{−/−} and TRAF6^{−/−} macrophages. Indeed, IKKβ-deficient macrophages generated by crossing *Ikkβ*^{F/F} mice¹² with mice expressing Cre recombinase from the IFN-inducible MX1 promoter¹³, were defective in NF-κB activation (data not shown) and underwent apoptosis upon incubation with LPS, LTA or TNF-α, even without p38 inhibition (Fig. 2e). IKKβ-deficient macrophages were also more susceptible to poly(IC)-induced apoptosis. Deletion of IKKβ in macrophages did not affect p38 expression (Fig. 2e) or activation (data not shown).

Although TLR4 activation results in TNF-α production, the apoptosis observed in TLR4-activated and p38-inhibited macrophages was not prevented by ablating type I TNF-α receptor (unpublished results). Furthermore, unlike apoptosis induced by TNF-α, caspase-8 was not activated during LPS-induced apoptosis

(Supplementary Fig. 1a). Yet, cleavage of caspases 3, 6, 7 and 9 and cytochrome *c* release were readily observed (Supplementary Fig. 1b; data not shown), and apoptosis was inhibited by a pan-caspase inhibitor (Supplementary Fig. 1c). Hence, inhibition of p38 or IKK unleashes the ability of TLR4 to deliver a cell death signal through the mitochondrial-dependent pathway¹¹.

Another protein involved in TLR signalling is the dsRNA-responsive kinase PKR¹⁴. PKR was suggested to mediate apoptosis in fibroblasts in response to viral infection and inflammatory cytokines¹⁵. However, PKR also activates IKK and NF- κ B¹⁶ and thereby suppresses apoptosis. To determine the role of PKR in macrophage apoptosis, we examined its regulation and found it to be rapidly activated by either LPS or poly(IC) (Fig. 3a). Activation of PKR by either stimulus depended on TRIF (Fig. 3b). Ablation of PKR did not affect p38 or IKK activation in response to LPS (Fig. 3c) and with the exception of decreased IFN β (Fig. 3d) or inducible NO synthase (iNOS; Fig. 3e) expression, PKR^{-/-} BMDMs did not exhibit reduced induction of numerous NF- κ B target genes, including those coding for anti-apoptotic proteins, such as c-IAP2, c-FLIP, A1a, A20, and Gadd45 β (Fig. 3d). PKR^{-/-} BMDMs also exhibited

defective STAT1 phosphorylation in response to LPS (Fig. 3f), which depends on autocrine production of type I IFNs because it was not observed in macrophages deficient in type I IFN receptor (IFNRI) (Supplementary Fig. 2).

Most importantly, PKR^{-/-} macrophages did not undergo apop-

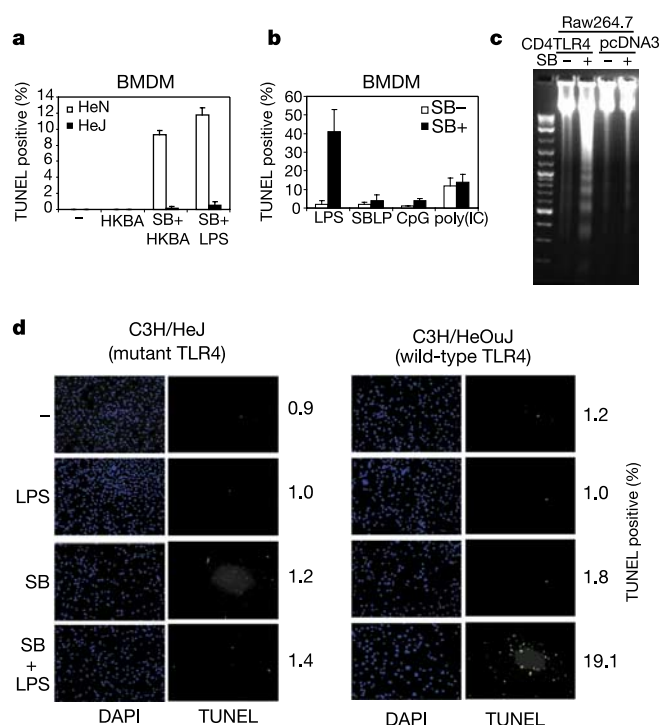


Figure 1 Heat-killed *B. anthracis* (HKBA) and LPS induce macrophage apoptosis through TLR4. **a**, HKBA induces macrophage apoptosis in a TLR4-dependent manner. BMDMs from C3H/HeN (*Tlr4* wild-type) or C3H/HeJ (*Tlr4* mutant) mice were incubated with HKBA or LPS (100 ng ml⁻¹) with or without p38 inhibitor (SB202190, 10 μ M). After 18 h, apoptosis was scored by TUNEL staining. Results in this and all similar experiments were repeated several times and one representative done in triplicates is shown. Values represent averages $\pm$ s.d. **b**, TLR4 agonists induce apoptosis in the presence of a p38 inhibitor. C57BL/6 BMDMs were treated with different TLR agonists: LPS, synthetic bacterial lipopeptide (1 μ g ml⁻¹), synthetic CpG-containing DNA (1 μ M), or poly(IC) (10 μ g ml⁻¹), with or without SB202190. **c**, TLR4 cytoplasmic domain transduces an apoptotic signal. RAW264.7 cells were transfected with a vector encoding a CD4-hTLR4 fusion protein, or an empty vector (pcDNA3). After 24 h, transfectants were incubated with or without SB202190, genomic DNA was isolated after 18 h and analysed by agarose gel electrophoresis for a nucleosomal ladder indicative of apoptosis. **d**, TLR4 is required for induction of apoptosis. C3H/HeJ or C3H/HeOuJ BMDMs were incubated with or without LPS in the presence or absence of SB202190 for 18 h and analysed by DAPI (blue) and TUNEL (green) staining. The percentage of TUNEL-positive cells is shown on the right.

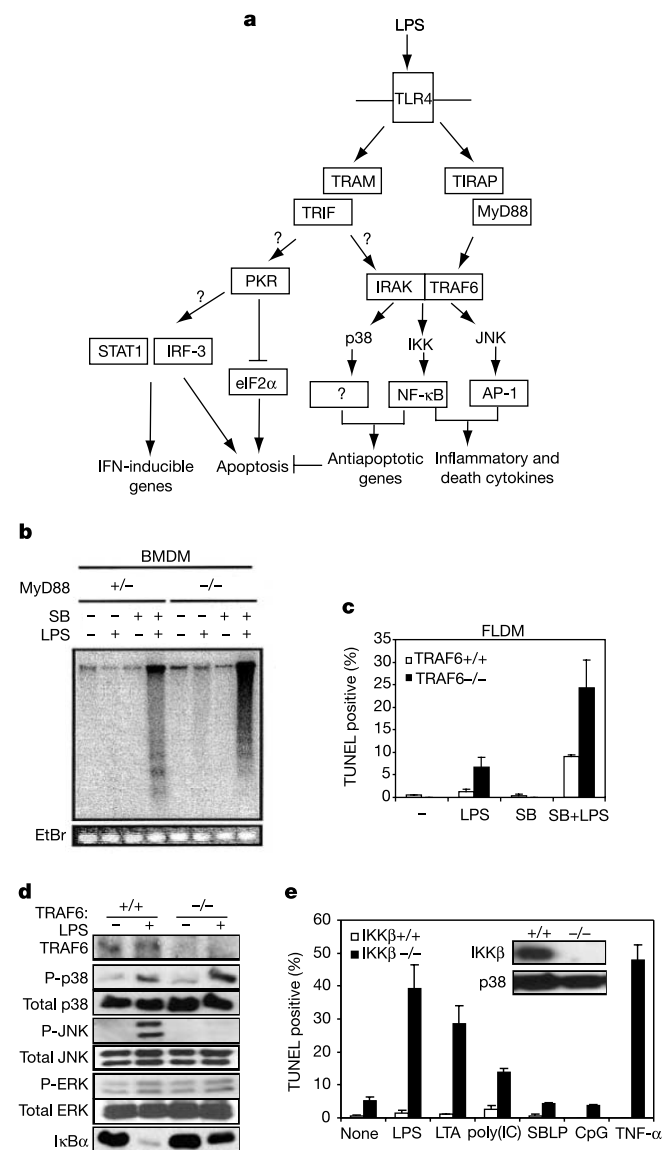


Figure 2 Role of effector molecules in TLR4-induced apoptosis. **a**, A diagram of currently known TLR4-stimulated signalling pathways in macrophages. Question marks denote yet-to-be established connections. **b**, Role of MyD88. MyD88^{+/+} and MyD88^{-/-} BMDMs were incubated with or without LPS in the presence or absence of SB202190. After 18 h, genomic DNA was isolated and end-labelled with [α -P³²]-dATP and Taq polymerase, followed by agarose gel electrophoresis and autoradiography. Gel loading was examined by EtBr staining. **c**, Role of TRAF6. Fetal-liver-derived TRAF6^{+/+} and TRAF6^{-/-} macrophages (FLDMs) were stimulated as above, and apoptotic cell death was quantified by TUNEL staining. **d**, Characterization of LPS signalling. TRAF6^{+/+} and TRAF6^{-/-} FLDMs were untreated or treated with LPS. After 20 min, cell lysates were prepared and analysed by immunoblotting with antibodies specific to TRAF6, Ikb α , different MAPKs and their phosphorylated forms. **e**, Role of IKK β . BMDMs from *Ikk β ^{+/+}* (*Ikk β ^{+/+}*) or *MX1Cre-Ikk β ^{F/F}* (*Ikk β ^{-/-}*) mice were untreated or treated with LPS, LTA from *B. subtilis* (10 μ g ml⁻¹), poly(IC) (10 μ g ml⁻¹), synthetic bacterial lipopeptide, SBLP (1 μ g ml⁻¹), CpG-containing DNA (1 μ M), or mouse TNF- α (10 ng ml⁻¹). After 12 h, apoptotic cell death was quantified by TUNEL staining. Inset, macrophage lysates were analysed by immunoblotting with antibodies specific to IKK β and p38.

apoptosis upon incubation with LPS and either SB202190 or LT (Fig. 4a) or after prolonged incubation with LPS alone (Supplementary Fig. 3a). PKR^{-/-} macrophages were, however, fully sensitive to other pro-apoptotic stimuli, such as doxorubicin or 15-deoxy- $\Delta^{12,14}$ prostaglandin J₂ (Supplementary Fig. 3b). PKR activation contributes to induction of type I IFNs, such as IFN β , which can further increase its expression¹⁷. Type I IFNs sensitize myeloid cells to LPS, LT or bacterial-induced apoptosis¹⁸. Whereas IFN β alone did not induce apoptosis in BMDMs, it rendered them susceptible to LPS-induced apoptosis even without SB202190 (Fig. 4b). IFN β also potentiated the apoptosis of wild-type BMDMs exposed to LPS + SB202190, but PKR^{-/-} BMDMs were resistant to these effects of IFN β (Fig. 4b). Nevertheless, BMDMs from IFNRI^{-/-}

mice¹⁹ still underwent apoptosis when incubated with LPS and SB202190 (Fig. 4c) and exhibited normal PKR activation by LPS (Supplementary Fig. 4). Thus, although PKR contributes to IFN β production, which can potentiate macrophage apoptosis, IFN β signalling itself is not essential for TLR4-induced PKR activation or macrophage apoptosis.

PKR is activated by poly(IC), which induces macrophage apoptosis even without p38 inhibition (Fig. 1b). Activated PKR phosphorylates eukaryotic translation initiation factor 2 α (eIF2 α) at serine 51 and thereby inhibits protein synthesis²⁰. Although the kinase activity of PKR is not required for NF- κ B activation, it is needed for induction of apoptosis (Supplementary Fig. 5). In addition, macrophages are very sensitive to protein synthesis inhibition (Supplementary Fig. 6). Hence, PKR activation may cause macrophage apoptosis by inhibiting synthesis of anti-apop-

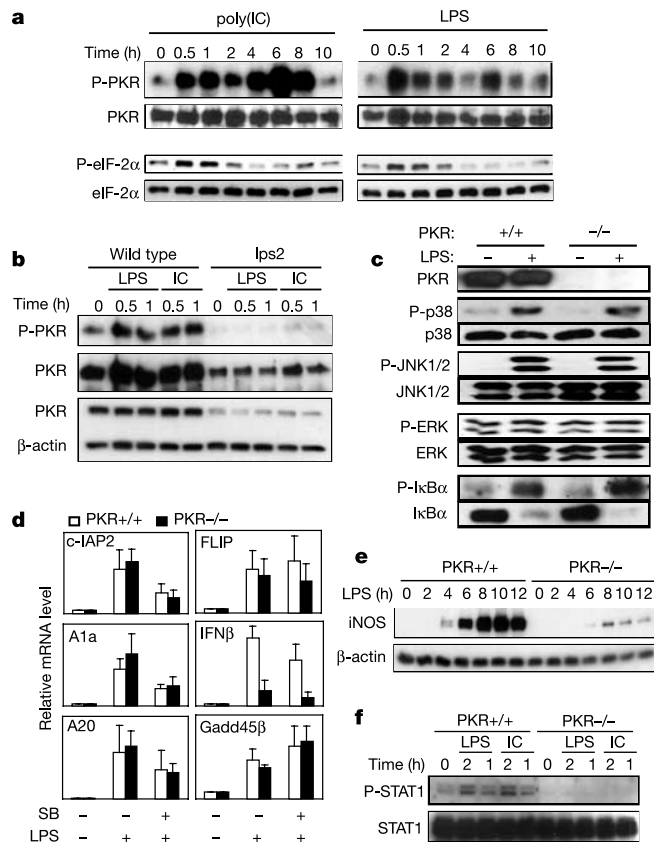


Figure 3 PKR is necessary for LPS-induced interferon signalling pathway in macrophages. **a**, Activation of PKR by LPS. BMDMs were stimulated with LPS (100 ng ml⁻¹) or poly(IC) (10 μg ml⁻¹). At the indicated time points cells were lysed and PKR activation was monitored by autophosphorylation. Gel loading was controlled by immunoblotting for PKR. The lysates were monitored for eIF-2 α phosphorylation by immunoblotting with antibodies specific for phosphorylated eIF-2 α (P-eIF-2 α) and total eIF-2 α . **b**, PKR acts downstream of TRIF. Wild-type and *lps2* (TRIF-deficient) BMDMs were stimulated with LPS or poly(IC). PKR activation was monitored by autophosphorylation. The same lysates were examined for PKR and β -actin content by immunoblotting. **c**, Normal MAPK and IKK activation in PKR^{-/-} macrophages. PKR^{+/+} and PKR^{-/-} BMDMs were left unstimulated or stimulated with LPS. After 20 min, cell lysates were prepared and immunoblotted with antibodies to different MAPKs or I κ B α and their phosphorylated forms. **d**, PKR is required for IFN β induction but is dispensable for induction of anti-apoptotic genes. PKR^{+/+} and PKR^{-/-} BMDMs were incubated with or without LPS in the absence or presence of SB202190. After 4 h, total cellular RNA was isolated, and relative gene expression was determined by real-time PCR. The results are averages of three separate experiments normalized to the level of cyclophilin mRNA. **e**, **f**, PKR is required for iNOS induction (**e**) and STAT1 phosphorylation (**f**). PKR^{+/+} and PKR^{-/-} BMDMs were incubated with LPS or poly(IC). At the indicated time points, cell lysates were prepared and iNOS expression and STAT1 phosphorylation were examined by immunoblotting.

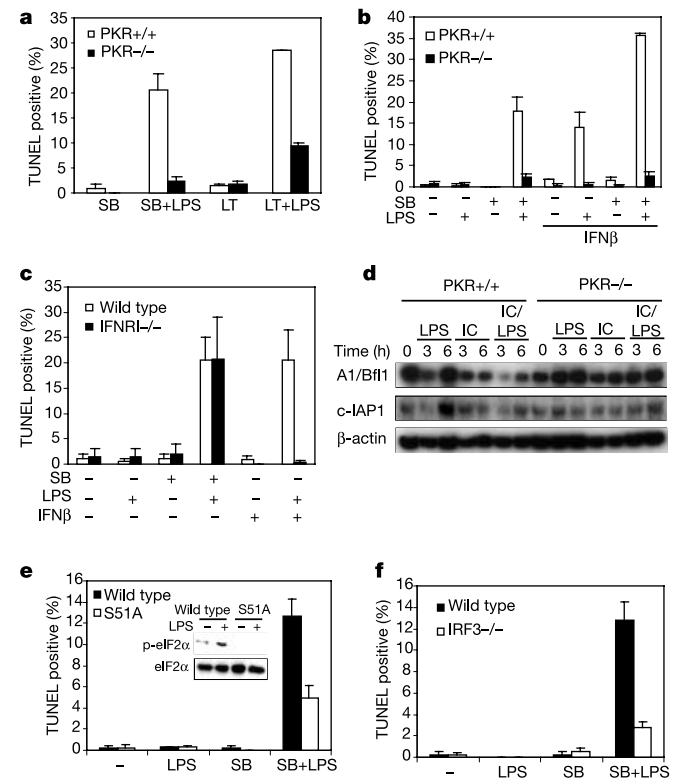


Figure 4 PKR is required for LPS-induced macrophage apoptosis. **a**, PKR-deficient BMDMs are resistant to LPS-induced apoptosis. PKR^{+/+} and PKR^{-/-} BMDMs were left unstimulated or stimulated with either LPS in the presence or absence of either SB202190 or LT (500 ng ml⁻¹ LF and 2.5 μg ml⁻¹ protective antigen, PA), and the extent of apoptosis was determined after 18 h. **b**, IFN β sensitizes BMDMs to LPS-induced apoptosis. PKR^{+/+} and PKR^{-/-} BMDMs were incubated with or without LPS in the absence or presence of SB202190 or IFN β (1,000 U ml⁻¹) for 18 h and the extent of apoptosis was determined. **c**, Type I IFN signalling is not required for LPS-induced macrophage apoptosis. BMDMs from IFNRI^{+/+} or IFNRI^{-/-} mice were incubated with or without LPS in the presence or absence of SB202190 and IFN β for 18 h and the extent of apoptosis was determined. **d**, PKR activation inhibits A1/Bfl1 expression. PKR^{+/+} and PKR^{-/-} BMDMs were transfected with or without poly(IC) using Lipofectamine. After 6 h, LPS was added and the levels of A1/Bfl1 and cIAP-1 were examined by immunoblotting. **e**, eIF2 α phosphorylation is required for induction of macrophage apoptosis. BMDMs derived from lethally irradiated mice reconstituted with fetal liver stem cells from either wild-type or *eIF2 α (S51A)* mice²² were incubated with LPS with or without SB202190 and the extent of apoptosis was analysed. The inset shows the absence of eIF2 α phosphorylation in knockin macrophages. **f**, IRF3 is required for induction of macrophage apoptosis. BMDMs from wild-type or IRF3^{-/-} mice were incubated with or without LPS in the presence or absence of SB202190 for 18 h and the extent of apoptosis was determined.

otic proteins. To enhance PKR activation and bypass TLR3, which may activate anti-apoptotic pathways, we introduced poly(IC) into macrophages by transfection²¹. This resulted in higher levels of PKR activity and caused a higher level of apoptosis (data not shown). Transfection of wild-type BMDMs with poly(IC) followed by LPS treatment inhibited accumulation of A1/Bfl1, an anti-apoptotic member of the Bcl2 family known to inhibit LPS-induced apoptosis of neutrophils¹¹, and to a lesser extent c-IAP1. The same treatment of PKR^{-/-} BMDMs did not reduce the level of either protein (Fig. 4d).

To examine directly the role of eIF2 α phosphorylation in macrophage apoptosis, we transplanted fetal liver haematopoietic progenitors from *eIF2 α (S51A)* knockin mice, which die shortly after birth, to lethally irradiated wild-type mice, in order to obtain macrophages that express an eIF2 α variant in which serine 51, the major PKR phosphorylation site, was replaced with an alanine²². *eIF2 α (S51A)* BMDMs were considerably less sensitive than wild-type BMDMs to apoptosis caused by incubation with LPS and SB202190 (Fig. 4f). The residual apoptotic response in *eIF2 α (S51A)* macrophages suggested the existence of another PKR-dependent pro-apoptotic pathway. It was proposed that interferon response factor 3 (IRF3), a transcription factor activated by dsRNA, is an important mediator of virus-induced apoptosis²³. We found that BMDMs from IRF3^{-/-} mice²⁴ showed increased resistance to LPS + SB202190 (Fig. 4f).

The role of TLR4 and PKR in apoptosis induced by live pathogenic bacteria was investigated using BMDMs infected with *B. anthracis*, *Yersinia* and *Salmonella*. PKR^{-/-} macrophages showed markedly reduced levels of apoptosis compared with PKR^{+/+} cells (Fig. 5a, b). The result with live anthrax bacilli was similar to the one obtained with LT regarding the PKR dependence of apoptosis, but did not require addition of LPS (compare Fig. 5a to Fig. 4a). Pathogens that induce macrophage apoptosis activate TLR4 through cell wall components, but apoptosis also requires a specific contribution from the bacteria. *Yersinia* spp. injects YopJ, an inhibitor of MAPK and IKK activation²⁵, into the host cell cyto-

plasm. As expected, a *yopJ* mutant of *Yersinia pseudotuberculosis* did not induce apoptosis in PKR^{+/+} macrophages (data not shown).

For *Salmonella* infections, PKR-dependent macrophage apoptosis requires the SPI2 locus (data not shown), which is responsible for translocation of bacterial virulence proteins from the phagosome into the macrophage cytoplasm²⁶. However, the *Salmonella* SipB protein, a caspase 1 activator²⁶, was dispensable for PKR-dependent apoptosis (data not shown), a finding consistent with the distinct mechanism of SipB-mediated cell death that differs from classical apoptosis²⁶. Consistent with the results described above, *eIF2(S51A)* and IRF3^{-/-} macrophages were also less susceptible to *Salmonella*-induced apoptosis (Fig. 5c, d). BMDMs from TLR4-deficient mice exhibited a dramatically reduced apoptotic response after infection with *B. anthracis* (Supplementary Fig. 7) and as shown in Fig. 1a, the *Tlr4* mutation in the C3H/HeJ strain prevented macrophage apoptosis by heat-killed *B. anthracis*. The apoptotic response to *Salmonella* and *Yersinia* was also considerably reduced in BMDMs from C3H/HeJ mice (Fig. 5e). Thus, the TLR4 to PKR pathway is crucial for macrophage apoptosis elicited by both Gram-positive and Gram-negative pathogens.

Microbial-induced macrophage apoptosis may represent a major mechanism allowing pathogenic bacteria to avoid detection and destruction by the innate immune system³. Virulence factors used by certain pathogens to dismantle host defences were identified and in some cases shown to act through inhibition of anti-apoptotic signalling pathways^{4,25,26}. The results described above shed further light on this phenomenon and identify what appears to be a general mechanism used by three different bacterial pathogens, *B. anthracis*, *Yersinia* and *Salmonella*, to specifically kill activated macrophages. We find that macrophage apoptosis by either Gram-positive (*B. anthracis*) or Gram-negative (*Yersinia*, *Salmonella*) pathogens requires activation via TLR4. Curiously, however, TLR4 is not a typical death receptor with death domains that cause caspase-8 activation. In fact, TLR4 engagement results in activation of both anti-apoptotic and pro-apoptotic signalling pathways. Normally, the anti-apoptotic pathways, which depend on MyD88 and TIRAP/

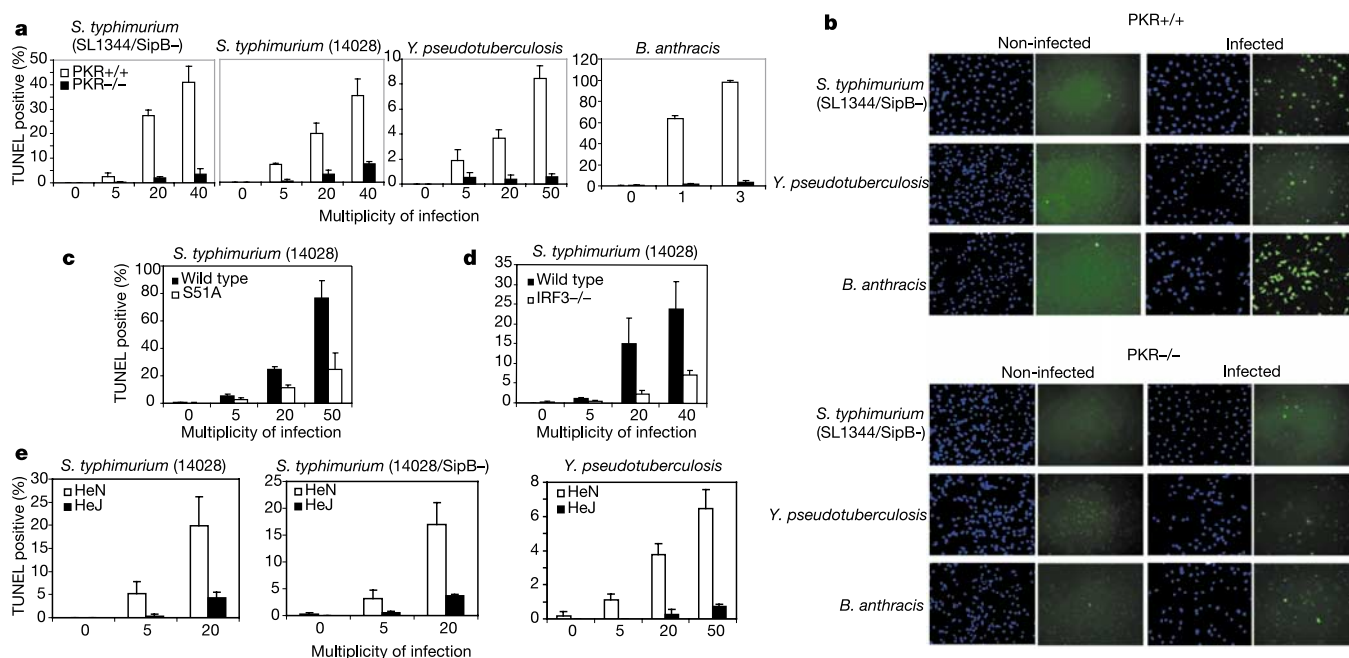


Figure 5 PKR-deficient macrophages are resistant to pathogen-induced apoptosis.

a, PKR^{+/+} and PKR^{-/-} BMDMs were infected with *S. typhimurium* (SL1344/SipB⁻), *S. typhimurium* (14028), *Y. pseudotuberculosis* or *B. anthracis* at the indicated multiplicity of infection. Apoptotic TUNEL-positive cells were scored 18 h post-infection. **b**, Representative DAPI (blue) and TUNEL (green) staining of PKR^{+/+} and PKR^{-/-} BMDMs

18 h post-infection with the indicated pathogens. **c**, **d**, wild-type and *eIF2 α (S51A)* (**c**), or IRF3^{-/-} (**d**) BMDMs were infected with *S. typhimurium* (14028) as above and the extent of apoptosis was determined 18 h later. **e**, BMDMs from C3H/HeN (*Tlr4* wild-type) and C3H/HeJ (*Tlr4* mutant) mice were infected with the indicated pathogens and the extent of apoptosis was determined after 18 h.

MAL, dominate, and exposure of macrophages to bacterial cell wall components, such as LPS, does not result in considerable cell death. However, at least two of the pathogens we examined produce specific virulence factors that inhibit survival pathways (p38, IKK/NF- κ B) and thereby tilt the balance in favour of cell killing.

We identified PKR as an essential component of the TLR4-triggered macrophage apoptosis pathway. On the basis of phenotypic and biochemical similarities in the behaviour of PKR- and TRIF- or TRAM-deficient macrophages^{7,8} and the requirement of TRIF for PKR activation, we propose that TLR4 activates PKR and triggers apoptosis through these newly described adaptors. Although PKR is an important contributor to antiviral defences under certain circumstances²⁷, its overall contribution to host defences in the case of bacterial infections has not been explored. The results described above suggest that PKR-inhibition may strongly augment macrophage-mediated anti-bacterial responses that do not depend on production of type I IFN and induction of IFN-responsive genes such as *iNOS*. The decreased expression of *iNOS*, a major inducer of vasodilation upon PKR inhibition, may also be taken advantage of for prevention of septic shock. □

Methods

Mice and bone-marrow-derived macrophages

To delete IKK β in macrophages, *Ikk β ^{+/F}* mice¹² were crossed with MX1-Cre mice (Jackson Laboratory). PKR^{-/-} mice²⁸, IFNRI^{-/-} (A129) and wild-type mice of the same genetic background (129/SvEv)¹⁹ were obtained from E. Raz. C3H/HeJ, C3H/HeOJ, C3H/HeN mice⁵ and bone marrow from lps2 mice⁷ were obtained from B. Beutler. TRAF6^{+/-}, MyD88^{+/-} and IRF3^{-/-} mice were received from J. Inoue, S. Akira and T. Taniguchi, respectively. C57BL/6J and C57BL/10ScCr (TLR4^{-/-}) mice were from the Jackson Laboratory. Unless otherwise mentioned, all knockout mice were of the C57BL/6 background, which is resistant to LT-induced necrosis. BMDMs were prepared and cultured as described⁴.

Analysis of gene expression and cell signalling

Total cellular RNA was prepared using TRIzol (Invitrogen), quantified by ultraviolet absorption and analysed by real-time polymerase chain reaction (PCR)⁴. Primer sequences are available upon request. All values were normalized to the level of cyclophilin messenger RNA expression. Whole-cell extracts were prepared and PKR activity was measured by autophosphorylation²⁰ after immunoprecipitation with anti-PKR antibody (Santa Cruz). PKR recovery was assessed by immunoblotting. Phosphorylation of eIF2 α was detected by immunoblotting with antibody against phosphorylated eIF2 α (Biosource). IKK and MAPK activation were measured as described⁴. Phosphorylation of STAT1 was monitored by immunoblotting with anti-phospho-STAT1 antibody (Cell Signalling).

Bacterial strains, macrophage infections and TUNEL assay

The wild-type *Salmonella typhimurium* strains used were SL1344 and 14028. *S. typhimurium* 14028 *ssaV* and *sipB* contain mutations in genes that code for components of the SPI2 type III protein secretion system and SipB, respectively. *Y. pseudotuberculosis* strains YP126 (wild type) and YP26 (YopJ⁻) were obtained from J. Bliska. *S. typhimurium* BMDM infection was as described²⁹, while *Y. pseudotuberculosis* infection was done as described³⁰ with slight modifications: BMDMs were infected with bacteria for 1 h, and then cultured in fresh medium containing gentamicin (20 μ g ml⁻¹) for another 18 h. The *B. anthracis* Sterne strain was grown overnight on BHI (brain-heart infusion) agar. A single colony was inoculated into BHI broth or RPMI medium + 10% FCS (endotoxin-free) in disposable tubes and grown with vigorous shaking to an OD₆₀₀ of 0.4. Bacteria were washed with PBS and resuspended in PBS. To prepare heat-killed *B. anthracis*, bacterial suspensions in PBS were heated to 65 °C for 30 min. Macrophage cultures were infected as indicated and incubated for 1 h at 37 °C in 5% CO₂/95% air. Gentamicin was added to a final concentration of 20 μ g ml⁻¹. After 20 h, the medium was removed and cells were fixed with 4% paraformaldehyde in PBS. TUNEL or DNA fragmentation assays were performed as described⁴.

Received 12 December 2003; accepted 2 February 2004; doi:10.1038/nature02405.

1. Aderem, A. & Ulevitch, R. J. Toll-like receptors in the induction of the innate immune response. *Nature* **406**, 782–787 (2000).

2. Medzhitov, R. Toll-like receptors and innate immunity. *Nature Rev. Immunol.* **1**, 135–145 (2001).
3. Weinrauch, Y. & Zychlinsky, A. The induction of apoptosis by bacterial pathogens. *Annu. Rev. Microbiol.* **53**, 155–187 (1999).
4. Park, J. M., Greden, F. R., Li, Z. W. & Karin, M. Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science* **297**, 2048–2051 (2002).
5. Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085–2088 (1998).
6. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. Jr A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
7. Hoebe, K. *et al.* Identification of Lps2 as a key transducer of MyD88-independent TIR signalling. *Nature* **424**, 743–748 (2003).
8. Yamamoto, M. *et al.* TRAM is specifically involved in the Toll-like receptor 4-mediated MyD88-independent signaling pathway. *Nature Immunol.* **4**, 1144–1150 (2003).
9. Kawai, T., Adachi, O., Ogawa, T., Takeda, K. & Akira, S. Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity* **11**, 115–122 (1999).
10. Naito, A. *et al.* Severe osteopetrosis, defective interleukin-1 signalling and lymph node organogenesis in TRAF6-deficient mice. *Genes Cells* **4**, 353–362 (1999).
11. Karin, M. & Lin, A. NF- κ B at the crossroads of life and death. *Nature Immunol.* **3**, 221–227 (2002).
12. Li, Z. W., Omori, S. A., Labuda, T., Karin, M. & Rickert, R. C. Ikk β is required for peripheral B cell survival and proliferation. *J. Immunol.* **170**, 4630–4637 (2003).
13. Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
14. Horng, T., Barton, G. M. & Medzhitov, R. TIRAP: an adapter molecule in the Toll signaling pathway. *Nature Immunol.* **2**, 835–841 (2001).
15. Kaufman, R. J. Double-stranded RNA-activated protein kinase mediated virus-induced apoptosis: a new role for an old actor. *Proc. Natl Acad. Sci. USA* **96**, 116935 (1999).
16. Chu, W. M. *et al.* JNK2 and IKK β are required for activating the innate response to viral infection. *Immunity* **11**, 721–731 (1999).
17. Samuel, C. E. Antiviral actions of interferons. *Clin. Microbiol. Rev.* **14**, 778–809 (2001).
18. Lehner, M., Felzmann, T., Clodi, K. & Holter, W. Type I interferons in combination with bacterial stimuli induce apoptosis of monocyte-derived dendritic cells. *Blood* **98**, 736–742 (2001).
19. Muller, U. *et al.* Functional role of type I and type II interferons in antiviral defense. *Science* **264**, 1918–1921 (1994).
20. Saelens, X., Kalai, M. & Vandenabeele, P. Translation inhibition in apoptosis: caspase-dependent PKR activation and eIF2- α phosphorylation. *J. Biol. Chem.* **276**, 41620–41628 (2001).
21. Diebold, S. S. *et al.* Viral infection switches non-plasmacytoid dendritic cells into high interferon producers. *Nature* **424**, 324–328 (2003).
22. Scheuner, D. *et al.* Translational control is required for the unfolded protein response and *in vivo* glucose homeostasis. *Mol. Cell* **7**, 1165–1176 (2001).
23. Heylbroeck, C. *et al.* The IRF-3 transcription factor mediates Sendai virus-induced apoptosis. *J. Virol.* **74**, 3781–3792 (2000).
24. Sato, M. *et al.* Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN- α / β gene induction. *Immunity* **13**, 539–548 (2000).
25. Orth, K. *et al.* Disruption of signaling by Yersinia effector YopJ, a ubiquitin-like protein protease. *Science* **290**, 1594–1597 (2000).
26. Rosenberger, C. M. & Finlay, B. B. Phagocyte sabotage: disruption of macrophage signalling by bacterial pathogens. *Nature Rev. Mol. Cell Biol.* **4**, 385–396 (2003).
27. Durbin, R. K., Mertz, S. E., Koromilas, A. E. & Durbin, J. E. PKR protection against intranasal vesicular stomatitis virus infection is mouse strain dependent. *Viral Immunol.* **15**, 41–51 (2002).
28. Yang, Y. L. *et al.* Deficient signaling in mice devoid of double-stranded RNA-dependent protein kinase. *EMBO J.* **14**, 6095–6106 (1995).
29. Browne, S. H., Lesnick, M. L. & Guiney, D. G. Genetic requirements for salmonella-induced cytopathology in human monocyte-derived macrophages. *Infect. Immun.* **70**, 7126–7135 (2002).
30. Zhang, Y. & Bliska, J. B. Role of Toll-like receptor signaling in the apoptotic response of macrophages to *Yersinia* infection. *Infect. Immun.* **71**, 1513–1519 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. Raz, V. Redecke, S. Akira, J. Inoue, T. Taniguchi and B. Beutler for various knockout mice and bone marrow, J. Bliska for *Yersinia* strains, N. Sonenberg and R. Medzhitov for gifts of plasmids and antibodies, and M. Delhase for technical assistance. L.-C.H., J.M.P., J.-L.L. and S.M. were supported by postdoctoral fellowships from the Cancer Research Institute, the Irvington Institute, the International Union Against Cancer and the Japanese Society for Promotion of Science, respectively. Work in the laboratories of M.K., D.G.G. and L.E. was supported by grants from the National Institutes of Health. M.K. is an American Cancer Society Research Professor.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.K. (karinoffice@ucsd.edu).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:
Nevin Bayouni (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel +49 89 54 90 57 11/-2

Fax +49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

Capping active volcanoes

A couple of years ago, I came across an unusual use for the term 'extinct volcanoes' (see *Naturejobs* 3; 24 January 2002). In the academic world, it seems, the phrase refers to tenured professors who have ceased to be productive and who simply occupy space and eat up resources, even though their passion has long since cooled.

But, as with most issues, there is a flip side. Some countries, notably in Europe, inadvertently cap 'active volcanoes' — professors of a 'certain age' who are still actively conducting research and publishing. Mandatory retirement for top scientists can actually promote brain drain at a time when Europe is trying to prevent it.

For example, when Alex Matter, who helped to develop one of the most promising cancer drugs, Gleevec (see page 348), reached 65, Swiss law meant that he had to retire as head of the oncology therapeutic area at the Novartis research department in Basel, Switzerland. Fortunately, Novartis found him work in Singapore, where he now runs the company's Institute for Tropical Diseases (see *Nature* 426, 588; 2003).

Frank Gannon, executive director of the European Molecular Biology Organization, recently argued that age discrimination in science should not be practised in Europe (F. Gannon *EMBO Rep.* 5, 221; 2004). He believes that if 79-year-old conductor Neville Marriner can still lead the orchestra of the Academy of St Martin in the Fields, then top scientists can still lead in research after they turn 65.

Both Gannon and I sympathize with the young scientists waiting for academic slots to open up when senior scientists retire. But we also wonder about the sanity of a system that blithely quenches brightly burning intellectual fires.

Paul Smaglik
Naturejobs editor



Contents

CAREERS & RECRUITMENT

Opportunities in
cancer research p348

CAREER VIEW

Scientists & Societies

An introduction to
biotechnology

Graduate Journal

Choosing a boss

Movers

Kim Nasmyth p350

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Growth industry

Drug companies are seeking help in their efforts to use the new understanding of cancer's complexities, says Ricki Lewis.

Discoveries in cell and molecular biology, new tools such as gene-expression profiling, and information from the human genome sequence are converging to brighten employment prospects in industrial cancer research. Two of the first cancer drugs to use these sources for 'rational' design — Herceptin, a protein-based treatment for breast cancer, and Gleevec, a small-molecule chemotherapeutic for some forms of leukaemia — are proving successful, and others such as Avastin and Erbitux are already following in their footsteps. This trend has created some bright spots in cancer R&D recruitment, even though drug discovery in general has been adversely affected by mergers, a few high-profile failures and a shaky US economy.

"Everyone is recognizing that there are going to be a lot of opportunities with all of the new information," says Jeff Hanke, vice-president for cancer research at AstraZeneca R&D Boston in Waltham, Massachusetts. Strong medical demand and a tremendous amount of new science makes this the time for a big push, he says. With that in mind, the company has taken on R&D staff in the United States and Britain during the past few years and built a new cancer facility in Waltham. And although opportunities are patchy, with some companies expanding cancer research and some not, there is currently a general upswing in hiring.

A GROWING MARKET

Signs of growth beyond AstraZeneca are clear. About 25% of biotech companies raising venture capital during the third quarter of 2003 listed cancer as their primary focus, according to online newsletter *VentureReporter*. And according to the Pharmaceutical Research and Manufacturers of America, 402 medicines were in development for cancer in 2002, up from 215 in 1996.

One catalyst is the Food and Drug Administration's accelerated drug-approval timetable: its evaluation of Novartis's Gleevec took just three months compared with the standard 10–12 months. But it is the staggering profits of early successes that attract most attention.

In 2003, \$1.9 billion of the \$3.3 billion in revenue collected by Genentech in South San Francisco came from oncology products, mostly the monoclonal antibody-based drugs Rituxan, used to treat non-Hodgkin's lymphoma, and Herceptin.

Successes of individual drugs are having "a ripple effect throughout the industry", says Vishva Dixit, vice-president, molecular oncology, at Genentech. And those ripples are reaching the job market. "Oncology is the most active area in pharmaceutical R&D. The job market in cancer research is pretty good at the moment," says

Steve Projan, assistant vice-president at Wyeth Oncology Discovery in Pearl River, New York.

Cancer is more of a problem in ageing populations, and the number of cases in the United States is expected to double by 2050, according to the National Cancer Institute (NCI) — unless there are changes in risk, improved diagnosis or new treatments. Although cancer cases are on the rise, so too are pharmaceutical success stories, and these have encouraged private funding. "The sometimes spectacular results with Herceptin have reinforced the idea that there is much to hope for as more biologics are brought to market," says Paul Schimmel, a molecular biologist at the Scripps Research Institute in La Jolla, California. "Much private capital has gone into developing biologics, and this has generated tens of thousands of jobs — many of them high-paying."

Increased understanding of carcinogenesis is providing new ideas, but market demands and profit potentials are also driving forces. "The ability to fund cancer research fluctuates and is greatly dependent on both internal and external economic factors, which drive hiring," says John Barnes, senior director of human resources at ILEX Oncology in San Antonio, Texas.

SIGNAL STRENGTH

Gleevec led the way in exploiting signal-transduction pathways to treat cancer — it blocks tyrosine kinase enzymes that can help to trigger out-of-control cell division. "Kinase inhibitors can be given safely and with minimal toxicity and have dramatic clinical effects," says Charles Sawyers, a Howard Hughes Medical Institute professor at the Jonsson Cancer Center at the University of California, Los Angeles, and part of the drug's discovery team.

The approach of targeting key signalling molecules is now spreading beyond the tyrosine kinases. For example, AstraZeneca's Iressa inhibits epidermal growth factor and is used to treat non-small-cell lung cancer. Thios Pharmaceuticals in Emeryville, California, focuses on the biochemical pathways that add sulphur atoms to molecules to activate their signalling capabilities. "As our efforts in sulphation targets progress, we will definitely be looking to expand the number of scientists devoted to oncology," says Thios president Bruce Hironaka. A cell biologist with experience in signal transduction could segue nicely into industry, as those pathways are the inspiration for new drug targets.

Subcellular systems other than signalling suggest new targets by revealing new points of vulnerability. The proteasome, for example, a structure that straightens out or dismantles misfolded proteins, is targeted by Velcade, a drug made by Millennium Pharmaceuticals in Cambridge, Massachusetts, that is used against refractory multiple myeloma.





Hands on: industrial research into cancer therapeutics has received a boost from the recent flood of genome data.

Another new avenue in cancer research is to combine drugs. Wyeth's Mylotarg, for instance, links an antibody to a chemotherapeutic, and homes in on CD33 receptors on acute myeloid leukaemia cells. Expertise in biochemistry, cell biology and immunology is required to develop such a drug.

BACK TO THE FUTURE IN HIRING

The bread and butter of biology — considering entire organisms — is returning to prominence, after decades of focus on the cellular and molecular. New institutes and academic departments devoted to 'systems biology' or 'integrative biology' reflect this trend, and it is important in the industrial cancer-research lab, too (see *Nature* **427**, 568–569; 2004). A good example of the kind of person Genentech seeks is a researcher who can create a transgenic or knockout animal model of a human cancer, and can then image the animal as the disease progresses and it reacts to treatment with a drug candidate, says Dixit.

"Ten years ago, cloners were valuable," Dixit says. "Now we are back to people who have biological insight coupled to expertise in technologies such as transgenics or DNA microarrays."

Because cancer perturbs cells in many ways, expertise in molecular and cell biology is always in demand, as is an ability to wade through vast amounts of data. Not all companies require incoming staff to have a PhD. At AstraZeneca, new employees range from a BSc or MSc with little experience to researchers with decades of experience, says Hanke. At Genentech, employees without a PhD can become a research associate, with pay comparable to that of an associate professor.

In general, as company size increases, focus narrows and salary increases, says Barnes. In other words, a researcher at a smaller company must have broader

skills, because there are fewer people to take a drug from concept to reality. "Opportunities are probably better in smaller biotech companies, as young researchers can take over responsibility faster," says Walter von Horstig, chief financial officer of Europroteome in Hennigsdorf, Germany. "Salaries are probably lower. A typical starting salary for a researcher with postdoc experience is about €45,000–50,000 (US\$55,000–60,000)."

But those in big drug companies can still be creative, says Ronny Mosston, senior director for staffing at Millennium Pharmaceuticals. "We nurture independent thinking and innovation." Wyeth also encourages researchers to develop their own ideas, says Philip Frost, vice-president at Wyeth Oncology Discovery.

PERSONALIZED PRESCRIBING

In the past, cancers were classified by body part — breast, lung or colon, for example. Gene-expression profiling is changing that picture by splitting cancers into molecularly defined subtypes. This approach will eventually reach the clinic, although the transition from bench to bedside is expected to be slow.

Hanke says that the new strategy will not fragment the market, but reclassify it according to drug response, based on gene-expression patterns. Molecular targeting of drugs will increase efficacy, so that patients will use them for a long time, adds Frost. And increasing profits may translate into more jobs.

With cancer increasingly being seen as treatable, the future for cancer research is bright. "There is exciting science, opportunity, emerging successful clinical trials, and you know you are working in an area of such great medical need," says Hanke. "All of these coalesce and make cancer research a very attractive area."

Ricki Lewis is a freelance science writer in Scotia, New York.

GRADUATE JOURNAL

Choosing a boss

Professor X is never in the office, doesn't respond to e-mails, meets you once every six months and is guaranteed not to hassle. Dr Y, meanwhile, requires daily updates and detailed progress reports, is always close at hand and knows your thesis research better than you do.

Choosing the right supervisor is crucial, yet many students underestimate the importance of their decision. Finding the right person can make or break a PhD.

The right kind of boss is someone whose style matches the way you like to be managed. You may prefer to be left alone to do your own thing, whereas your boss wants to guide you every step of the way. That doesn't make them bad, just the wrong style for you.

Research is rarely done in a vacuum, and students usually interact with a wide spectrum of people. My PhD has taught me as much about myself and other people as it has about physics. If you have the right people to work with, half the battle is won.

When your relationship with your boss is flexible enough to suit both of you, you can get the most out of your research. A good supervisor is like a best friend who helps you to ride the PhD roller-coaster safely. So take time to find the right combination for you; it'll be worth it in the end.

Amber Jenkins is a graduate student in particle physics at Imperial College, London, doing thesis research at Fermilab in Batavia, Illinois.

SCIENTISTS & SOCIETIES

An introduction to biotechnology

Being a graduate student or postdoctoral fellow — even if you're fortunate enough to work in a good lab and on a fun project — can be an isolating experience. You live and breathe academic science, and as a result you can develop a limited perspective on the world and on your career.

To broaden my horizons, I joined Yale's Biotechnology Student Interest Group, which aims to provide a big picture of the biotech industry, beyond just the science. The group's mission is to bring together graduate students and postdocs in the biomedical sciences, business, medicine, public health and law, for educational and career-related opportunities relating to the biotech and drug industries. This somewhat dry-sounding

vision doesn't quite capture the exciting and dynamic character of the organization. It has about 750 members, including postdocs, graduate students, professional students in law and business as well as many members of the small but dynamic biotechnology community centred around New Haven, Connecticut.

Its speaker series helps to broaden members' perspective of the life-sciences industry by including academics engaged in biotech, patent attorneys, industry chief executives, venture capitalists and investment bankers. For example, at the annual 'biotechnology reception' last October, Steven Holtzman, chief executive of drug-discovery firm Infinity Pharmaceuticals, dropped his self-described "standard PowerPoint presentation". Instead, he drew on his Rhodes scholar training in philosophy, and his

experience as a former member of the National Bioethics Advisory Commission, to present an analysis of the ethical and philosophical implications of stem-cell and cloning technologies.

But the organization does more than listen to speakers. It also provides free consulting services to local biotech companies. And it has an equity-analysis group that evaluates firms for investment from financial, legal and scientific perspectives.

Being involved with the organization has not only widened my scientific perspective — it has helped me to make a career change. Once I have finished my postdoc this summer, I plan to enrol as an MBA student at Yale, and train to manage a biotech company.

Eric Anderson is a postdoc at Yale University and a co-president of the university's Biotechnology Student Interest Group.

♦ www.yale.edu/biotech

MOVERS

Kim Nasmyth, Whitley chair of biochemistry, University of Oxford, UK



In 1988, Kim Nasmyth moved from Cambridge, UK, to the recently established Research Institute of Molecular Pathology (IMP) in Vienna, Austria. He knew almost no German, and had few connections in the close-knit Austrian research network. But his enthusiasm for mountaineering and skiing — and the fortuitous proximity of an old climbing partner — helped him to adapt.

During the 1990s, Vienna moved from the scientific periphery to the heart of a flourishing central European life-sciences community. And Nasmyth quickly rose through the ranks to be one of its leaders, becoming director of the IMP in 1997.

Climbing and skiing in the Austrian Alps has been a satisfying counterweight to Nasmyth's triple role as scientist, institute director and research manager. But Nasmyth now feels ready for a change. In early 2006, he will take over the Whitley chair of biochemistry at the University of Oxford, UK.

The move will mark a return to a completely academic environment for Nasmyth. Although the IMP is a basic-research institution, it is funded by German drug firm Boehringer Ingelheim. Nasmyth hasn't regretted his marriage to an industrial partner. The dowry was generous — professionalism, clearly defined hierarchies and minimal bureaucracy for raising cash. "There have been many advantages," says Nasmyth. "But research outside universities tends to be extremely focused, which is not always the best thing."

Nasmyth will take over the Whitley chair from Edwin Southern. He knows he

has big boots to fill — but he feels well prepared. The IMP served as a springboard for a number of successful scientific careers, and he hopes to repeat this success in Oxford.

At the IMP, Nasmyth helped to nurture scientific talent. "The trick is to hire them in a steady trickle, and not too many people too quickly," he says. At Oxford, he will take the same approach, and will generate an environment where peer pressure, not funding pressure, is the driving force towards excellence. "Expectations must always be high," he says. "But I'm keen that young researchers are still capable of taking pleasure in other people's success."

Neither the IMP nor Austria will be easy to leave, says Nasmyth. "I'll definitely miss the mountains," he says. But he has already prepared for his absence — he has bought a farmhouse in southern Austria as a holiday base for future climbs.

CV

1997–: Director, Research Institute of Molecular Pathology, Vienna, Austria

1988–97: Senior scientist, Research

Institute of Molecular Pathology, Vienna, Austria

1982–87: Staff member, Medical Research Council Laboratory of Molecular Biology, Cambridge, UK